

НАУЧНО • ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ

Серия 2. ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ И СИСТЕМЫ
ЕЖЕМЕСЯЧНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ СБОРНИК

Издается с 1961 г.

№ 12

Москва 2015

ОБЩИЙ РАЗДЕЛ

УДК 004.65 : [54.03/.04 : 620.22]

А.О. Еркимбаев, В.Ю. Зицерман, Г.А. Кобзев, А.В. Косинов

Связывание онтологий с базами данных по свойствам веществ и материалов*

Рассмотрены возможности нового подхода, интегрирующего потенциал онтологии и традиционной базы данных, применительно к задачам хранения и распространения данных по свойствам веществ и материалов. Представлен обзор нескольких технологий проектирования и интеграции научных баз данных, ориентированных на их объединение с онтологиями. Показано, что использование онтологии предметной области обеспечивает стандартизацию терминов и понятий, исключая их неоднозначность и ошибочную интерпретацию, унификацию семантики, согласованную в экспертном сообществе, распространение знаний, сопровождаемое возможностью логического вывода и автоматизацией поиска. В качестве примеров использования онтологии рассмотрены: одна из наиболее масштабных баз данных в области химии ChEBI; несколько систем хранения и интеграции данных по свойствам материалов; авторский проект базы данных по теплофизическим свойствам индивидуальных веществ.

Ключевые слова: численные данные, базы данных, онтология, систематизация данных

* Работа выполнена при поддержке РФФИ – проект № 13-07-00218

ВВЕДЕНИЕ

Накопление и систематизация фактографических данных о свойствах веществ и материалов издавна является одним из основных направлений деятельности в естественнонаучных и инженерных дисциплинах: физической химии, кристаллографии, материаловедении и др. [1–3]. Широкое использование для этих задач электронных ресурсов (баз данных, вычислительных систем, Web-порталов и др.) при постоянно возрастающем объеме их использования позволяет отнести эту сферу научной деятельности к так называемой *e-Science*, которую отличает не столько компьютеризация исследования, сколько преимущественная ориентация на работу с данными, включая их получение, систематизацию, хранение и распространение для коллективной работы. Особенности *e-Science* позволяют говорить о новом этапе в научной методологии, *4-й парадигме* [4], дополняющей стандартную триаду: теория, эксперимент и моделирование.

Ключевым (хотя далеко не единственным) элементом при работе с данными о свойствах является база данных (БД). Если на определенном этапе переход к БД от многотомных справочников обеспечил заметные достижения в создании и поддержке информационных ресурсов, то со временем резкий рост объемов данных при многообразии их форматов и моделей привел к тому, что действующая инфраструктура оказалась неспособной обеспечить эффективную организацию рабочего процесса [5].

Автономность БД с жесткой фиксацией используемых терминов и логических структур практически исключает возможность беспрепятственного обмена данными без активного участия человека-эксперта. В информационном сообществе в качестве главного ограничителя возможностей БД называют так называемую *потерю семантики* в ходе проектирования БД, когда принятая концептуальная модель реализуется в виде физической модели. По сути, это означает, что каждая из БД имеет собственный понятийный аппарат, который (без специальных инструкций) остается неизвестным для пользователя. Чтобы построить правильный запрос, пользователь любой из БД должен знать названия сущностей и их атрибутов, как минимум – принятые названия веществ и их физических свойств. При этом любые тематически родственные БД с неизбежностью имеют различия в концептуальной схеме.

Применительно к свойствам вещества первым шагом к преодолению неоднородности БД была бы стандартизация названий путем использования общедоступных и одинаково трактуемых индексов, например **CAS Register Number**. По сути, это проблема повторяет на новом этапе ту же проблему стандартизации, которая еще до компьютерной эры заняла важное место в международном сотрудничестве в физике, химии, кристаллографии и др. Под эгидой IUPAC (Международный союз по чистой и прикладной химии) и CODATA (комиссия по численным данным в науке и технологии) разработаны многочисленные рекомендации для обозначений, терминологии, химической номенклатуры, единиц

измерения и пр. Повсеместный переход к компьютерным средствам заметно обостряет проблему, поскольку встает вопрос не только одинаковой трактовки человеком, но и совместимости данных с множеством приложений и компьютерных средств поиска. Поэтому даже такой простейший аспект, как разноречивость в названии веществ, обусловленный богатой синонимией, может стать непреодолимым барьером при поиске.

Проблема интеграции ресурсов и, прежде всего, БД достаточно давно обсуждается как в компьютерном сообществе, так и среди экспертов по систематизации научных данных [1, 3, 5, 6]. Один из очевидных подходов – выработка принятой в сообществе логической схемы, обеспечивающей стандартизацию в именовании веществ, свойств, состояний и прочих аспектов. В качестве интегрирующей модели стала использоваться модель, основанная на стандартах XML, долгое время служившего наиболее используемым форматом при обмене структурированной информацией – между программами, между людьми, между компьютером и человеком [7]. XML, или расширяемый язык разметки (*eXtensible Markup Language*) создан для описания данных и концентрирует свое внимание на их **сущности**, что принципиально отличает его от языка HTML, созданного для отображения, т.е. показа данных. Существенно, что XML – это свободный и расширяемый язык, где разработчик имеет возможность предлагать свои теги, используя которые можно создать структуру документа в полной аналогии с таблицами реляционной БД. Тот факт, что данные хранятся в обычных текстовых файлах, позволяет в технологии XML использовать программно и аппаратно независимые решения. Обмен данными путем их преобразования в XML-формат позволяет упростить связь несовместимых сред и делает данные доступными для различных программ, благодаря чему он стал языком, применяемым для обмена информацией через Интернет, а сами XML-документы своеобразными хранилищами данных. Стандарт под названием XML Schema может быть принят как стандарт структуры данных в определенной профессиональной среде.

Разумеется, XML-технология была подхвачена в естественнонаучных дисциплинах для целей обмена и интеграции численных данных. Во множестве дисциплин появились собственные версии XML [8] со своими словарями, средствами поддержки в виде настраиваемых браузеров и программ, реализующих графические представления, вычислительные сервисы и пр. Они позволяют работать с данными самой разной структуры и содержания, например, полученными в молекулярной динамике (MODML), вычислительной нейрофизиологии (NeuroML), биоинформатике (BSML) и даже *спелеологии* (Cave Survey Markup Language). Широкую известность и распространение получили языки для описания химической информации (CML) [9] и математических выражений (MathML) [10].

Для стандартизации обмена данными в термодинамике и материаловедении были разработаны и широко анонсированы языки ThermoML и MatML. Проект ThermoML [11], поддержанный IUPAC, в качестве основной цели имел стандартизацию форм

хранения и обмена теплофизическими данными для более чем 120 свойств при различных формах представления и различном статусе данных (экспериментальные, расчетные, справочные). Близкие по сути задачи призван решать язык MatML [12, 13], с той разницей, что объектами являются не вещества с известной стехиометрией, а материалы, свойства которых существенно зависят от технологии изготовления, факторов внешнего воздействия и т.п. Возможности и ограничения в применении обоих языков достаточно подробно описаны в нашей работе [6].

К этой же тематике, связанной с физическими свойствами, относится стандарт XSAMS: XML Schema for Atomic, Molecular and Solid Data [14], созданный для совместимости БД, содержащих параметры атомных и молекулярных процессов, а также характеристик взаимодействия частиц с поверхностью твердого тела. Стандарт предлагает физикам универсальную структуру, но только для обмена данными, без контроля их достоверности. Процесс описывается путем ссылки на начальное и конечное состояние, а сами данные представлены либо в табличной форме, либо в виде параметров аппроксимирующих выражений. Поскольку считаются важными источник и происхождение данных, схема накладывает жесткие требования на информацию по источникам данных и методам их генерации. Полное описание структуры XSAMS-документов можно найти в руководстве [15].

Постепенно, однако, в информационном сообществе пришло осознание недостатков, присущих XML-технологиям, особенно при стандартизации данных в науке. XML-схема обеспечивает синтаксис и структуру данных, но не их семантику (т.е. смысл), необходимую при гибком отображении словарей и схем данных. В итоге для тех же целей роль XML-схем заняли онтологии, формализующие в закодированном виде структуру знаний. Подробнее их возможности в организации научного знания описаны в следующем разделе. Здесь достаточно отметить, что размещение в сети онтологии, закодированной на языке OWL (Web Ontology Language), позволяет каждому из понятий предметно-ориентированного словаря присвоить уникальный идентификатор, так называемый URI (Uniform Resource Identifier), на который могут ссылаться автономные ресурсы, обеспечивая его единую трактовку.

Ориентация на онтологии легла в основу концепции **онтологически основанных БД**, которая предусматривает семантическое отображение структур данных, потенциально сводящее локальные БД в единое пространство данных при согласовании их терминологии и логической структуры. Доступность семантики при компьютерной обработке данных – не единственное преимущество онтологий при их сопоставлении с традиционными БД или жестким XML-стандартом. К достоинствам онтологического подхода относятся: легкость запроса информации при общедоступной семантике; возможности интеграции ресурсов; эффективные средства проектирования БД.

Осознание огромного потенциала, заложенного в онтологиях для систематизации и распространения научного знания, стимулировало активную деятельность по разработке предметно-ориентированных онтологий для формализации отдельных сегментов многих научных дисциплин: медико-биологических, химии, наук о Земле и др. Появились онтологии для наук о материалах [16–18], а также наноструктур [19]. Параллельно в сфере информационных технологий разработаны новые продукты, полученные гибридизацией БД и онтологий. Механизм работы всех этих продуктов использует некоторое взаимное отображение между концептуальной схемой БД и онтологией.

В целом традиционные БД и онтологии по своим возможностям для систематизации данных оказались взаимно дополнительны. БД способны к структурированию и хранению больших объемов данных при высокой эффективности их поиска. С одной стороны, онтологии позволяют устранить их главные дефекты – отсутствие общепонятной семантики, жесткую, неспособную к эволюции структуру. С другой стороны, сама онтология совместно с экземплярами данных не способна заменить БД из-за падения эффективности поиска с ростом объема данных.

Цель настоящей работы – выявить возможности и преимущества новых технологий, использующих связку «онтология–БД» при организации научных данных по свойствам вещества. Онтология обеспечит для предметной области, определяемой кругом объектов и набором свойств, терминологический словарь, соотношения между понятиями и структуру данных, а ее связка с БД – возможности хранения при эффективном поиске больших объемов данных.

ОНТОЛОГИИ И БД – ВОЗМОЖНОСТИ СОВМЕСТНОГО ИСПОЛЬЗОВАНИЯ

В последние годы, концепция **онтологий**, исходя из возникшая как философская категория, перешла в сферу информатики, где она определяет структуру для формальной спецификации понятий и отношений в определенной области знаний. Наиболее простым и точным ее определением считается то, которое дал Грубер [20]: «Онтология есть спецификация концептуализации», т.е. стандартизованное и закодированное представление знаний в определенной предметной области. Кодирование или запись на машинно-интерпретируемом языке (например, OWL) обеспечивает принципиальную возможность восприятия и поиска данных программными агентами. Главное назначение онтологии в том, что они добавляют к данным семантику (смысл, правильное толкование) и соотношения, что в совокупности описывает «знание» с возможностью его машинной интерпретации.

При разработке онтологии используют четыре типа составляющих элементов: понятия/концепты, соотношения, экземпляры и аксиомы. Концепты – это классы сущностей или объектов в определенной области знания, например, вещества, свойства, состояния и т.п. Отдельные классы связаны соотношениями – либо типа класс–субкласс (таксономия), либо ассоциативными соотношениями, которые определяют свойства или роли отдельных понятий. Общие понятия конкретизируют экземпляры, так что онтология

вместе с экземплярами, представленная на OWL, выполняет функцию записи данных, примерно так же, как XML-документ (см. Введение). Наконец, аксиомы устанавливают присущие данной области ограничения значений отдельных понятий или экземпляров. Например, при определении свойств молекул аксиомы могут выделить класс радикалов, по отношению к которым лишено смысла их отнесение к твердой фазе. Тем самым аксиомы позволяют достаточно формально представить многочисленные «реалии» предметной области.

Оценить богатый потенциал, заложенный в онтологиях для хранения и распространения данных, можно, сопоставляя их возможности с БД [21]. При кажущемся сходстве решаемых задач между ними имеются глубокие различия. Онтология служит для распространения информации, определяя на формальном языке концепции и соотношения, которые представляют содержание и структуру предметной области. В то время как концептуальная схема БД, определяя все понятия и структуру данных, служит только для тех целей, что реализует конкретная БД. Распространение информации посредством онтологии проводится согласованным образом, т.е. передаваемая ею структура данных является общедоступной и одинаково трактуемой в определенном сообществе. Все члены сообщества могут использовать онтологию и имеют доступ к информации.

Простой пример дает единое соглашение об именовании веществ, принятое согласно онтологии ChEBI (Chemical Entities of Biological Interest Ontology) [19], которая присваивает каждому из низкомолекулярных веществ уникальный идентификатор, и для согласованной номенклатуры необходима ссылка на эту онтологию и соответствующий веществу ID. Так, ID, присвоенный метану, имеет вид CHEBI:16183, а ссылка в любой из БД на уникальный URI www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:16183 обеспечивает

единую (и доступную для компьютера) трактовку этого термина. Аналогично можно обеспечить ссылки и на другие понятия, представленные в онтологии: названия свойств, физических состояний, единиц измерения и т.п.

Более полный перечень ключевых различий между БД и онтологией приведен в табл. 1 [21]. Наряду с семантикой, которая теряется в ходе проектирования БД, в таблице указаны некоторые дополнительные признаки. В частности, БД немыслима без содержания в виде данных, в то время как в онтологии наличие экземпляров не обязательно (*instances optional*). С другой стороны, онтология немыслима без таксономии классов (хотя и не сводится к ней), а в БД таксономия как структурный компонент отсутствует. В итоге формальный характер онтологии позволяет реализовать машинные выводы и рассуждения, что совершенно не предусмотрено в концептуальной схеме БД.

Применительно к тематике «свойства веществ и материалов» особую роль играет возможность на уровне онтологий поддерживать эволюцию схемы данных, связанную с расширением круга объектов и появлением новых, ранее неизвестных понятий (последний пункт в табл. 1). Как отмечалось в наших работах [1, 2, 22], вся практика систематизации данных по свойствам показывает невозможность поддерживать их логическую схему для широкой совокупности веществ, не вступая в противоречие с объективной потребностью передать их специфику и описания, принятые в разных коллективах. Хорошим примером служит возникшая необходимость перестройки стандарта ThermoML при переходе к задачам представления данных для свойств биоматериалов (ферментов и протеинов) и равновесий в растворах [23]. Другой пример эволюции схемы данных связан с постепенным включением наноструктур в химические БД, что с необходимостью ведет к изменению правил идентификации [1, 19].

Таблица 1

БД и онтологии – характерные признаки и отличия

Концептуальная схема БД	Онтология
Определяет структуру БД на формальном языке	Определяет набор концепций и соотношений, которые представляют содержание и структуру предметной области на формальном языке
Фокусируется на данных	Фокусируется на смысле
Сущности	Классы
Атрибуты	Соотношения
Ограничения	Аксиомы
Нет таксономии	Таксономия – ключевой элемент
Данные – ключевой элемент	Экземпляры данных не обязательны
Семантика только в концептуальной схеме, спроектирована для человека, не эволюционирует с изменением БД и приложений	Семантика – ключевой элемент, доступный программной обработке
Схему трудно изменять и поддерживать	Потенциально легче изменять и поддерживать

Вывод о безусловном преимуществе онтологий был бы, однако, совершенно преждевременным. Помимо того, что БД обеспечивают наиболее популярный и эффективный способ хранения данных, их отличает высокая производительность при поиске и реализации сложных запросов, недостижимая для других архитектур. Поэтому, рассматривая роль онтологий в информационной инфраструктуре, исследователи выбрали в качестве основной стратегии не «вытеснение» БД, а создание своеобразной связки путем так называемого *database-to-ontology mapping*. Ее задача использовать преимущества обеих конструкций, в основном за счет соединения семантики с высокой производительностью при работе с данными. Анализ предложенных концепций и продуктов дан в ряде обзоров [24–26]. Авторы [26], исходя из того, какая из двух концепций (онтология или схема БД) может рассматриваться как ведущая, выделили две стратегии. Первая из них ориентирована на использование онтологий для усиления функциональности БД. Тогда БД рассматривается как ведущий элемент, усиленный связью с онтологией для целей: реализации запросов с использованием согласованной в сообществе специалистов семантики; облегчения проектирования БД; обеспечения интеграции нескольких БД. Вторая стратегия, напротив, рассматривает БД как возможный инструмент усиления или даже разработки онтологии.

В соответствии с двумя стратегиями можно выделить две концепции взаимного отображения БД и онтологии и два типа продуктов. Первой концепции соответствует БД, основанная на онтологии (**DBBO** – *database based on ontology*), когда смысл сущностей, записанных в БД, определен ссылкой на соответствующую онтологию. Для каждой таблицы или атрибута устанавливается связь с концепцией, включенной в онтологию, – см. приведенный выше пример со ссылкой на название вещества в онтологии **ChEBI**. Вторая концепция предполагает проектирование структуры, получившей название онтологии, основанной на БД (**OBDB**, *Ontology Based on DB*). Задача такой структуры разместить экземпляры онтологии в БД, чтобы обеспечить достаточную эффективность при управлении, т.е. загрузке, поиске и реализации сложных запросов. В основном эта концепция нацелена на решение задач Semantic Web, где требуется организовать хранение и управление архивами документов, записанных на языке **RDF** (*resource definition framework*), который используется для связывания тематически родственных документов в глобальной сети [27].

В настоящей статье нас будет интересовать первая стратегия, ориентированная на технологию БД. Система, построенная по принципу **DBBO**, включает: исходно принятую БД с заполняющими ее записями; онтологию для семантического индексирования БД; возможные ссылки на другие онтологии с целью расширения словаря; соотношения между каждым элементом БД и онтологическим понятием. Онтология в такой структуре лишь обеспечивает семантику, но не содержит экземпляров – их роль выполняют записи БД. По сравнению с традиционной архитектурой БД, система **DBBO** включает два дополни-

тельных уровня. Первый из них (онтологический) включает одну или несколько онтологий для определения понятий предметной области, причем эти определения никак не связаны с конкретными приложениями, для которых спроектирована БД. Второй уровень (концептуальный) представлен соотношениями, связывающими онтологический уровень с логической моделью данных, причем выбор концептов может быть использован для автоматической генерации логической модели БД.

Следует заметить, что выделение концепции **DBBO** (как и сама аббревиатура), по-видимому, принадлежит авторам [26], в то время как другие авторы [24, 25] неправомерно используют аббревиатуру **OBDB** вне зависимости от конкретной схемы взаимного отображения онтологии и БД. Работ, посвященных методам усиления БД за счет связи с онтологией, пока немного, хотя уже предложены конкретные алгоритмы отображения и языки запросов, расширяющие возможности традиционного языка SQL. Особый интерес представляет алгоритм **Round-Trip Engineering** («поездка в оба конца») [28], способный обеспечить согласование реляционной схемы БД и концептуальной схемы или онтологии в тех случаях, когда возникает потребность в поддержке их эволюции при расширении информационного содержания. Алгоритм позволяет синхронизировать концептуально-реляционное отображение с появлением новой семантики.

Усиление функциональности за счет онтологии возможно не только для уже существующей БД, но и на этапе ее проектирования. От проектировщика требуются хорошие знания о специфике предметной области, которые он должен формализовать в виде концептуальной схемы, используя различные инструменты типа **UML**. Для одной и той же предметной области схемы разных экспертов будут радикально отличаться, а сами БД становятся гетерогенными, по сути, недоступными для машинно-реализуемой интеграции. Подход, при котором в основу проекта БД положена определенная онтология, заметно ускоряет процесс разработки и снимает ряд возникающих ограничений. В частности, концептуальные схемы, предлагаемые разными экспертами, в известной степени унифицируются, поскольку базируются на единой модели области. Кроме того, уже на этапе проектирования снимается проблема семантики, поскольку каждый элемент данных получается отображением онтологического понятий, имеющих точное определение и смысл.

Наряду с проектированием БД, использование онтологий открывает возможности в интеграции БД, ранее созданных без ссылок на онтологии. Методы интеграции издавна занимают важное место в технологии БД [3, 6, 7]. Под интеграцией в общем случае понимается их физическое или виртуальное соединение в одном месте, что позволяет использовать общий интерфейс и единую систему запросов. В ходе интеграции производится объединение схем и самих данных. Онтология предоставляет естественный и достаточно эффективный способ объединения схем, позволяя преодолеть проблемы с отсутствием семантики в исходных БД. Пользователь может формули-

ровать в терминах онтологии запросы, которые будут обращены к «подключенным» к онтологии БД. Тем самым онтология будет играть роль эффективного посредника между пользователем и данными.

Подобный подход получил название *доступ к данным, основанный на онтологиях*, или **OBDA** (*ontology based data access*). В **ODBA**-системе источники могут рассматриваться как единая БД. Хотя в целом такой подход к интеграции БД является относительно новым, уже сейчас, по мнению авторов [25, 26], можно указать несколько технологий, опробованных в разных проектах. Простейшая из них предполагает, что для набора БД принимается одна онтология как посредник между пользователем и данными. При этом сама онтология может включать в качестве модулей несколько специализированных онтологий. Существенно при этом, чтобы все объединяемые ресурсы примерно одинаково отражали специфику предметной области. Затруднения возникают, когда один из ресурсов отражает какой-то определенный аспект, не характерный для других. Кроме того, одна онтология может не справиться с задачей интеграции ресурсов, когда их содержание

расширяется так, что выходит за грань ранее принятой концептуализации. Альтернативой этой технологии служит использование множества онтологий, каждая из которых соответствует одной из подлежащих интеграции БД. В этом случае любые изменения или расширения, относящиеся к одному ресурсу, затрагивают только одну онтологию. Ключевым элементом технологии является процедура взаимного отображения онтологий (*alignment of ontologies*). Наконец, был предложен и промежуточный подход, когда каждому ресурсу ставится в соответствие одна онтология, но над ними располагается общий контролируемый словарь базисных терминов общей предметной области. Схематично различия трех технологий представлены на рис. 1.

В нашей работе [29] было показано, насколько эффективным оказывается онтология при проектировании БД по теплофизическим свойствам и интеграции данных из множества источников. Другие реализованные к настоящему времени примеры интеграции БД при посредстве онтологий по свойствам веществ рассмотрены в следующем разделе.

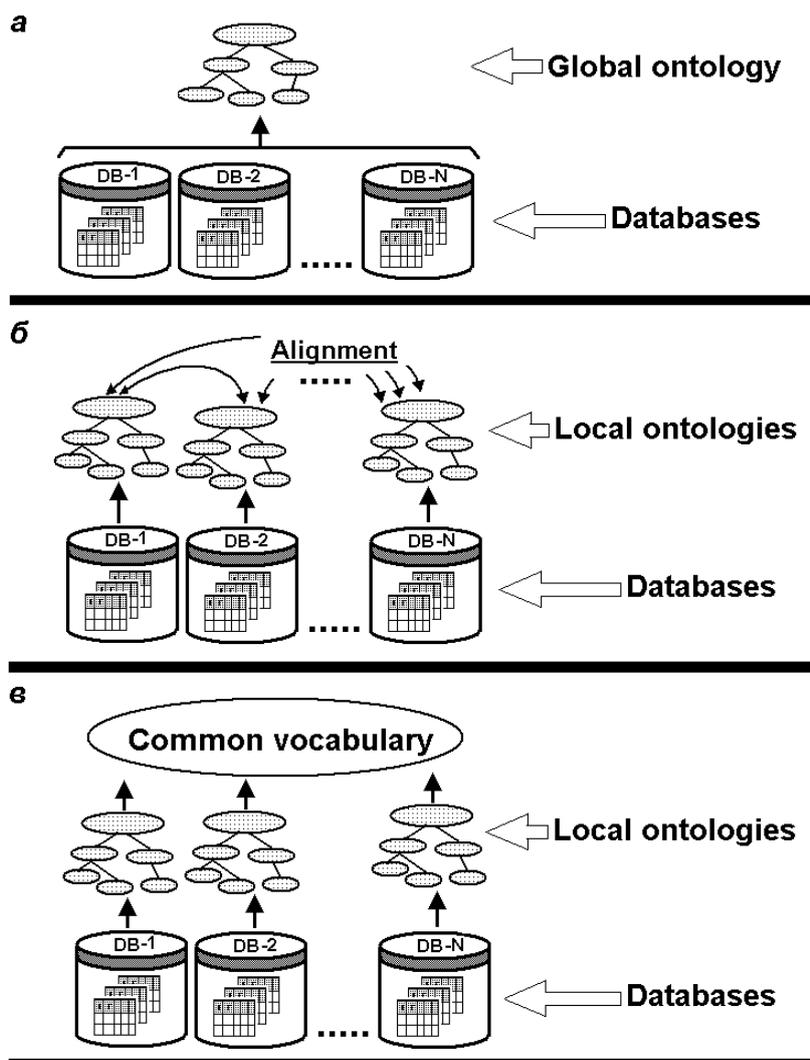


Рис. 1. Иллюстрация различий в технологиях **OBDA** (ontology based data access) [26]

ПРАКТИКА СВЯЗЫВАНИЯ ОНТОЛОГИЙ С БД ПО СВОЙСТВАМ ВЕЩЕСТВ И МАТЕРИАЛОВ

Один из наиболее впечатляющих примеров расширения функциональности БД за счет связывания с онтологией дает упомянутая выше БД химического содержания **ChEBI** [19], преимущественно ориентированная на низкомолекулярные соединения (исключены белки, нуклеиновые кислоты и пептиды). БД охватывает молекулярные сущности, их группы и классы. Первое из понятий относится к любым идентифицируемым по составу атомам, молекулам, ионам, наноструктурам и т.п., второе – к группам связанных атомов (или один атом) в составе молекулы (methyl, CH₃), а третье (класс) к совокупности молекул или групп, охваченных классификационным признаком, например, alkane (R-CH₃). Каждую сущность, а также группу и класс идентифицирует уникальный идентификатор **ChEBI ID** (например, **CHEBI:15377** для воды), свободный для цитирования в сети пользователем или программным агентом.

В записи для молекулярной сущности (рис. 2) указаны химические данные (определение, формула, масса, заряд), названия, как принятое в **ChEBI**, так и синонимы из других источников, структурная информация, кодированная в линейных нотациях (**InChI** и

SMILES), а также детализированная в **Molfile** (координаты атомов и матрица связности). Наряду с этим в записи представлены регистрационные номера в химических классификаторах (CAS, Beilstein и др.) и ссылки на другие БД с обширной информацией, например, термодинамической в БД NIST (webbook.nist.gov/chemistry/), или токсикологической в БД ChemIDplus (chem.sis.nlm.nih.gov/chemidplus/chemidlite.jsp).

Главным элементом онтологии, интегрированной с БД, является таксономия в виде ациклического графа, когда любое понятие может происходить от нескольких родительских понятий. Внизу рис. 2 показаны дочерние (incoming) и родительские (outgoing) классы, непосредственно связанные с веществом в записи (в данном случае **paracetamol**, **C8H9NO2**). На рис. 3 приведен фрагмент древовидной структуры, охватывающий всю таксономию, связанную с веществом, причем пользователь имеет возможность навигации с вызовом данных для любой сущности, представленной на «дереве». Обращение к URI с указанием онтологии и уникального ID непосредственно связывает внешний ресурс (или пользователя) с соответствующей записью, например URI www.ebi.ac.uk/chebi/searchId.do?chebiId=46195 связывает его с записью в БД для парацетамола.

The image shows a screenshot of the ChEBI entry for paracetamol. The entry includes a chemical structure, a list of properties, and search options. Annotations highlight specific features:

- Recommended ChEBI name:** paracetamol
- ChEBI ID:** CHEBI:46195
- Definition:** A member of the class of phenols that is 4-aminophenol in which one of the hydrogens attached to the amino group has been replaced by an acetyl group.
- Stars:** been manually annotated by the ChEBI Team.
- Secondary ChEBI IDs:** CHEBI:46191, CHEBI:2386
- Supplier Information:** eMolecules:474380, eMolecules:27677450, ZINC18274777
- Chemical structure searches:** Find compounds which contain this structure, Find compounds which resemble this structure, Take structure to the Advanced Search
- Additional chemical data:** Formula: C8H9NO2, Net Charge: 0, Average Mass: 151.16260, InChI: InChI=1S/C8H9NO2/c1-6(10)9-7-2-4-8(11)5-3-7/h2-5,11H,1H3,(H,9,10), InChIKey: RZVAJINKPMORJF-UHFFFAOYSA-N, SMILES: CC(=O)Nc1ccc(O)cc1

Рис. 2. Внешний вид типовой записи в БД ChEBI

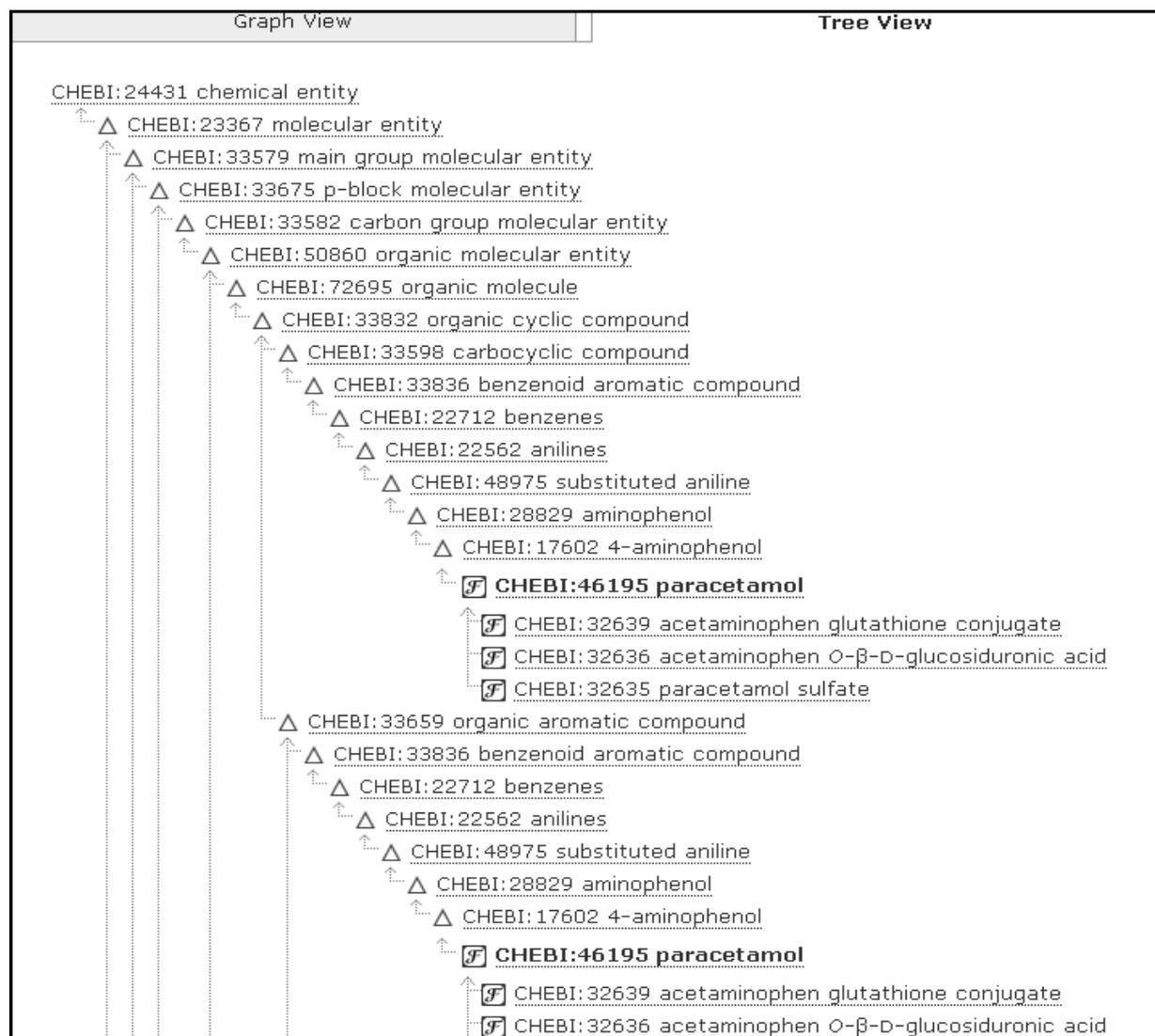


Рис. 3. Фрагмент «дерева», включающего сущность **ChEBI:46195** в общей таксономии

На том же рис. 3 (справа) приведен список ассоциативных свойств (соотношений), определяющих каждую сущность. Первые два (**is_a**, **has_part**) определяют простейшие виды логической связи, восемь остальных – фиксируют химический контекст в отношениях двух веществ, например, соотношение **has_functional_parent** (как видно из рис. 2) указывает, чем является *paracetamol* по отношению к другому веществу *paracetamol sulfate*. Особое значение в списке имеет последнее соотношение **has_role**, позволяя раскрыть множество аспектов в свойствах и приложении веществ. Оно связывает базовую субонтологию **Molecular structure**, в рамках которой построена таксономия веществ, с другой субонтологией **Role**, которая включает три класса верхнего уровня (**biological role**, **chemical role**, **application**), рис. 4. При этом **chemical role** классифицирует сущности по их «химически значимой» роли (кислота, основание, лиганд), **biological role** определяет их «биологически значимую» роль, а **application** дает классификацию по целевому использованию, например **drug**, **pesticide** и др. (рис. 4). Все сущности, относя-

щиеся к этой субонтологии, так же как и вещества, имеют уникальный ID, что позволяет связывать внешний ресурс с любой из сущностей в онтологии **role**. Пользователь может совершать навигацию по дереву, соответствующему ролевой онтологии, точно так же, как и для веществ, включенных в БД. Например, выделив дочерний субкласс **fuel additive** в классе **chemical role**, можно найти все вещества, относящиеся этому понятию (рис. 5) и выявить их характеристики и связи.

Таким образом, наличие онтологии открывает для пользователя или внешнего ресурса возможности, исходно отсутствующие в БД. В частности, по идентификатору **ChEBI** можно сделать ссылки на любые сущности БД, определить их логические и ролевые связи, встроенные в многоуровневые таксономии. В результате можно возложить на программного агента реализацию сложных запросов с поиском веществ по структуре, функциям, биологической роли и т.п. В целом это означает, что наряду с данными для конкретного вещества, **ChEBI** предоставляет фрагмент «знаний», характеризующих предметную область.

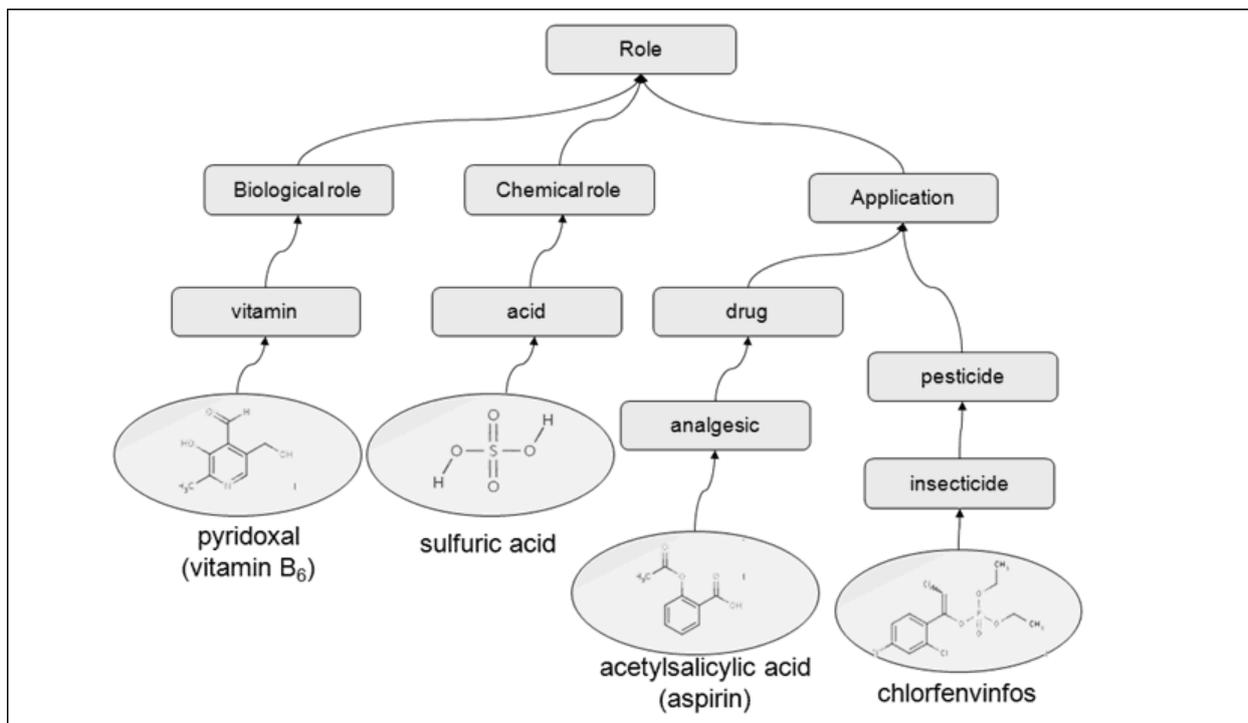


Рис. 4. Схематичное представление суб-онтологии **Role**

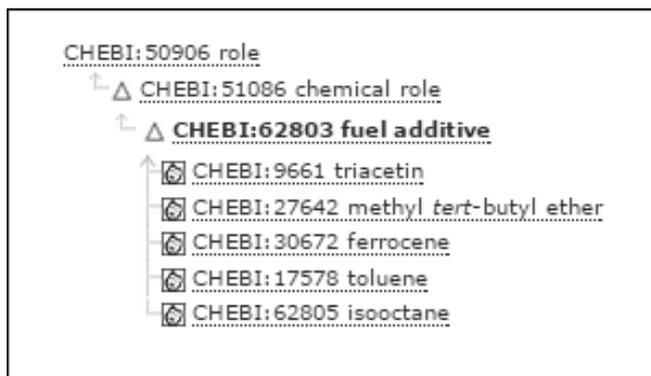


Рис. 5. Сущности, относящиеся к суб-классу **fuel additive** из суб-онтологии **chemical role**

Есть также несколько удачных примеров использования онтологий для интеграции данных в материаловедении, где многообразие типов данных и богатство словарей проявляются наиболее ярко. Среди таких примеров база знаний **PLINIUS**, оперирующая результатами исследований по свойствам керамик, онтологическое описание свойств ползучести конструкционных материалов, стандарт ISO 10303-235: “**Engineering properties for product design and verification**” с охватом информации о промышленных изделиях, наряду с традиционными материалами [6]. Стандарт предусматривает единую информационную модель для определения семантики и синтаксиса представления и единый словарь для определения смысла данных. В сравнении с ранее внедренным языком разметки MatML [13, 14], онтология способна обеспечить более высокий уровень стандартизации материаловедческих данных, формализующей определение свойств, методов обработки и использования.

В наиболее общем виде онтология по свойствам материалов предложена в работе [16]. Онтология, покрывающая предметную область, построена в виде семи субонтологий, распределенных по трем группам, рис. 6. Четыре базовые онтологии дают определения терминов, названий и словарей, представляющих основные концепции для каждой из областей. Каждая из онтологий основана на таксономии классов, представленных в словаре понятий. Базовые онтологии **Process** и **Environment** на рис. 6 дают описания соответственно методов производства и измерения и характеристик среды (состав атмосферы, температура, pH и т.п.). Сверх четырех базовых, в общую онтологию включена субонтология **Materials Information**, детализирующая сведения по конкретному объекту путем агрегирования других классов (вещество, свойство и т.п.). Используя базовые онтологии, эта онтология агрегирует все термины и концепции, характеризующие материал и конкретный образец, методы и условия измерения, критерии качества данных и пр. В частности, при построении онтологии **Unit Dimension** используется синтаксис MathML (версия, предназначенная для передачи формул), чтобы ввести операции, необходимые при согласовании различных единиц измерения.

Онтология, получившая название **Matinfo**, была размещена на портале [30] и встроена в общую схему Semantic Web. Основная задача, решаемая при посредстве разработанной онтологии [16], – обмен данными между тремя БД Японии: AIST (Advanced Industrial Science and Technology), NIMS (National Institute of Material Science) и широко используемой структурой **MatDB** [31], содержащей данные по тестированию материалов. Основной объем информации в указанных БД относится к тепловым и механическим свойствам. На рис. 7 показана часть структуры тепловых данных, представлен-

ных в двух БД (AIST, NIMS) и онтологии. Обмен данными реализован с использованием промежуточного уровня, занятого онтологией **Matinfo**. Стрелки на рисунке показывают взаимное отображение концепций онтологии и сущностей в БД.

Структура данных NIMS приспособлена к хранению экспериментальных данных, в соответствующих полях хранятся основные метаданные, **thermalConductivity** и **chemicalFormula**. Напротив, в БД AIST метаданные конкретизируются пользователем, в поле с именем **property** вводится текстовая строка, что позволяет дополнять или модифицировать имена

свойств при фиксированной схеме. Дополнительную сложность вносит то, что БД AIST способна хранить как скалярные, так и тензорные значения теплопроводности, характерные для монокристаллов или материалов с выраженной анизотропией строения. Эта же возможность закреплена и в онтологии, позволяющей отразить как скалярную величину, так и произвольную матрицу. Отображение полей каждой из БД производится посредством **XSLT** (*Extensible Stylesheet Language Transformations*) шаблонов, причем первичная подстройка осуществлялась в полуавтоматическом режиме.

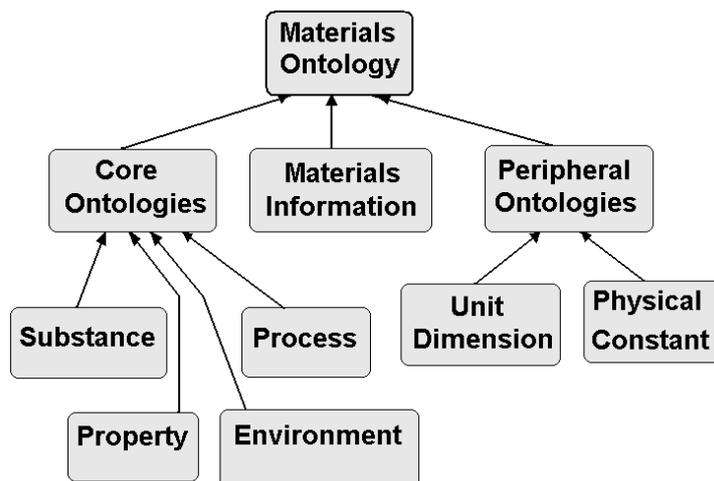


Рис. 6. Суб-онтологии, включенные в общую онтологию **Matinfo** по свойствам материалов [16]

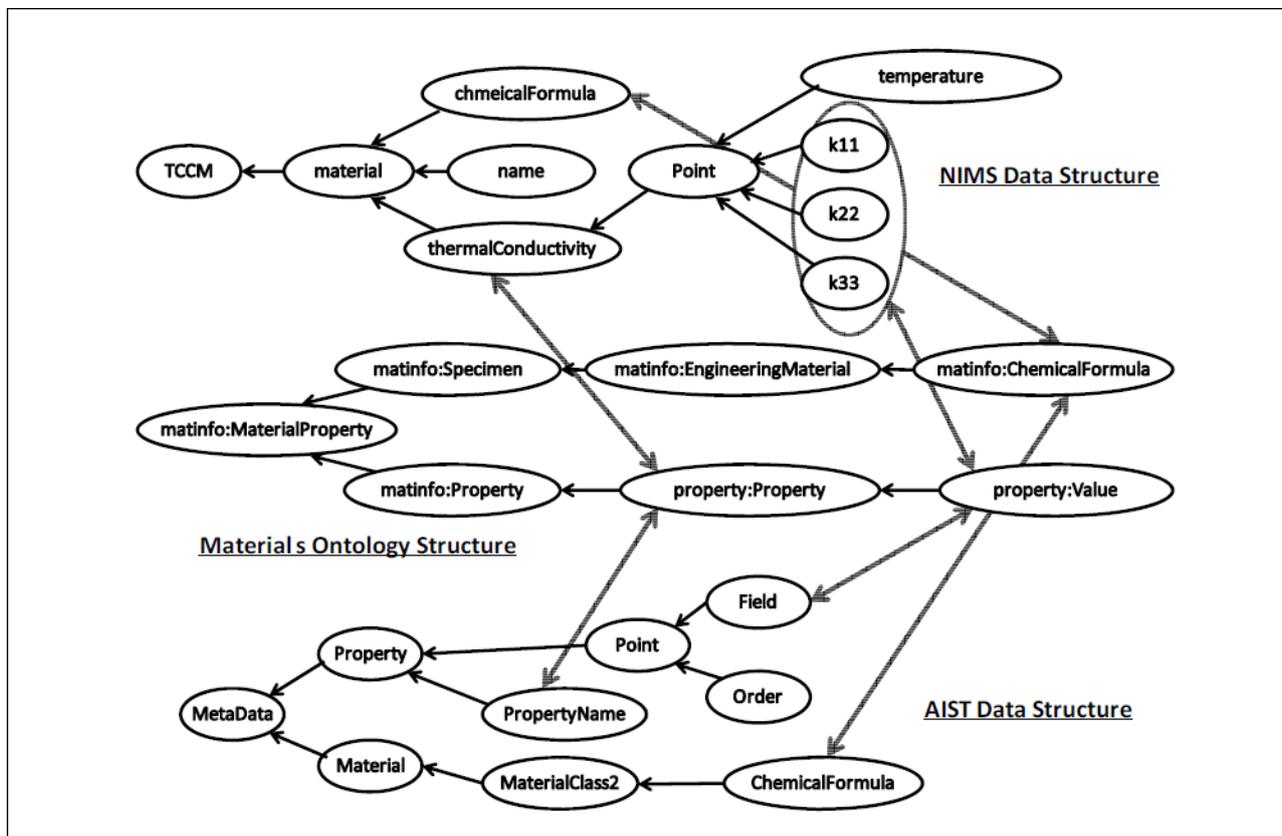


Рис. 7. Отображение схем данных для теплопроводности в БД AIST, NIMS и онтологии [16]

В соответствии с классификацией, рассмотренной в предыдущем разделе, использование онтологии **Matinfo** полностью соответствует концепции **OBDA** (доступ к данным, основанный на онтологии), причем здесь использована одна онтология, включающая в себя все понятия в науке о материалах, которые могут встретиться в интегрируемых источниках.

Аналогичные цели решает разработанная система **MatSeek** [17] с главной целью интеграции структурно и тематически разнородных БД. Концептуальную основу этой системы образует онтология **MatOnto**, которая аналогично **Matinfo** описывает базовые концепции и логические соотношения в науке о материалах. В ряде отношений, однако, эта онтология заметно отличается. Прежде всего, она нацелена на интеграцию данных из достаточно далеких предметных областей, которые объединены лишь тем, что используются на равных основаниях при исследовании. Например, при разработке топливных элементов в качестве интегрируемых ресурсов выбраны три БД: Inorganic Crystal Structure Database (ICSD), Ionic Radii database, и Phase Equilibria Diagrams (PED). Онтология, установленная в системе **MatSeek** на стороне сервера, записана на языке OWL и выступает посредником между пользователем и группой неоднородных БД.

Второе важное отличие, связанное с проведением логических рассуждений, состоит в активном привлечении внешних онтологий, в том числе онтологий так называемого верхнего уровня, аккумулирующих базовые научные концепции: процесс, гипотезу, эксперимент, активность и пр. Наконец, онтология встроена в достаточно сложную информационную структуру **MatSeek**, функционирующую как приложение в среде Semantic Web. **MatSeek** предоставляет Web-платформу для работы с БД и аналитическими средствами, позволяет формулировать основанные на онтологии динамические запросы к интегрируемым ресурсам, предлагает пользователю поисковый интерфейс в стиле Google. Ядром этой системы является онтология, выполняющая ряд задач: интеграцию разнородных структурно и тематически БД; представление информации о происхождении данных, включая описание источника, методов получения, неопределенности и т.п.; проведение формальных рассуждений с целью выделения нового знания. На рис. 8 показан верхний уровень онтологии с выделением основных классов и пояснением связей между концептами, относящимися к структуре и свойствам материалов. Как видно из рисунка, онтология имеет 5 классов верхнего уровня:

- **Property** – таксономия с 9 субклассами (механические, электрические и т.д.);
- **Family** – таксономия материалов с 6 субклассами (металлы, стекла, керамика, полимеры, гибридные материалы, эластомеры);
- **Process** – два субкласса (**Manufacturing** и **Measurement**);
- **Structure** – два субкласса (**Crystalline** и **Amorphous**);
- **Measurement Data** – таксономия результатов измерений, интегрирует 4 субкласса, охватывающих данные из разных источников (**Material Property-Data**, **PerformanceData**, **Modelling and Simulation-Data**, **CharacterizationData**)

Как уже говорилось, важным элементом, обогащающим выразительность онтологии **MatOnto**, является широкое привлечение концептов из других онтологий и словарей, например, из онтологии по единицам измерений, из академического предметного классификатора и др. Сверх того, понятия, актуальные для материаловедения, связаны с более абстрактными лингвистическими понятиями, присущими естественному языку. Для их внедрения использована онтология верхнего уровня **DOLCHE** (*Descriptive Ontology for Linguistic and Cognitive Engineering*), формулирующая такие абстракции как **Endurant** (thing, object), **Perdurant** (process, event), **Abstract**, **Quality**. Другая из привлеченных онтологий **EXPO** ориентирована на формализацию научного контента в любой из естественнонаучных дисциплин, ориентированных на эксперимент как основную методологию. Ключевые понятия **EXPO** определяют цели, объект эксперимента, выдвинутые гипотезы, результаты и т.п. Привлечение онтологий верхнего уровня обеспечивает расширяемость **MatOnto** с распространением на другие сегменты материаловедения.

Таким образом, потенциал онтологии **MatOnto** раскрывается за счет ее способности к интеграции неоднородных ресурсов, в том числе тематически удаленных, как, например, БД по ионным радиусам и фазовым диаграммам. По логике построения и решаемым задачам онтология, как и вся система **MatSeek**, соответствует концепции **OBDA**, когда онтология выступает как посредник между пользователем и набором разнородных БД.

Заслуживает внимания еще одна попытка использования онтологий в сфере материаловедения для доступа к опытным данным [18]. В отличие от рассмотренных проектов в духе **OBDA**, здесь предложена иная концепция, использующая трансформацию языка **MatML** в соответствующую онтологию. Эта трансформация преодолевает принципиальные ограничения **MatML** при обмене данными. Прежде всего, это древовидная структура документа, противоречащая практике запросов, использующих базовые термины предметной области. Кроме того, как отмечалось во Введении, XML схемы не описывают семантику, что и послужило основанием для отказа от **MatML** как средства интеграции данных в пользу онтологического моделирования. В то же время при переходе к онтологии в работе [18] предложено сохранить и интегрировать содержание **MatML**-документов – в какой-то степени по аналогии с реляционными БД. Этот подход предполагает набор правил для выделения из данной **MatML**-схемы соответствующей онтологии, включающей классы и соотношения, а также специальный алгоритм, преобразующий данные исходного документа в экземпляры онтологии. Построенной путем такой конверсии онтологии авторы [18] присвоили название **MatOWL**. При этом, поскольку **MatOWL** есть все-таки OWL-версия **MatML**-схемы, эта конверсия ограничена в части семантики. Поэтому предусмотрено расширение онтологии за счет дополнительных концепций, соотношений и аксиом – либо путем их прямого внедрения, либо путем отображения **MatOWL** на какую-либо из существующих онтологий в области науки о материалах.

Для конверсии самих данных, содержащихся в MatML-документах, в экземпляры онтологии предложена довольно сложная процедура, использующая промежуточную конструкцию – объектную модель **MatOO**. Модель имеет древовидную структуру как исходный MatML-документ, а также представлена в виде классов и соотношений, как должно быть в онтологии. Рис. 9 схематично представляет трансформацию данных в экземпляры онтологии в RDF-формате.

Таким образом, исходный материал в виде MatML-документов удастся конвертировать в типовой формат, предусмотренный Semantic Web для целей распространения в сети. Далее при большом объеме данных становится актуальной апелляция к технологиям **OBDB**, чья задача разместить экземпляры онтологии в БД, чтобы обеспечить достаточную эффективность при управлении, т.е. загрузке, поиске и реализации сложных запросов.

ПРОЕКТИРОВАНИЕ ТЕПЛОФИЗИЧЕСКОЙ БД НА ОСНОВЕ ОНТОЛОГИИ

Общая концепция **DBVO** была опробована ранее при создании БД по теплофизическим свойствам веществ [32]. Роль онтологии здесь особенно существенна,

поскольку, наряду с унификацией семантики, необходимо поддерживать эволюцию схемы данных и аксиомы, отражающие логические и математические ограничения, присущие данной области. Ключевое понятие – набор данных, включающий для одного вещества несколько констант и температурных функций, а также сведения о фазовом состоянии вещества, единицах измерений, неопределенности и источнике данных. Основные списки – веществ, свойств, фазовых состояний, единиц измерений и т.д. считаются открытыми, что позволяет в рамках онтологии поддерживать эволюцию схемы данных.

Концептуализация предметной области привела к выбору 12 базовых понятий, послуживших основой для построения соответствующих классов (табл. 2). Среди них группа из 4 базовых классов (вещества, состояния, свойства, численные данные), 6 вспомогательных классов и 2 класса, определяющих вычисляемые функции и аргументы. При этом класс **Functions** порождает 2 subclasses, определяющих значения свойств и выполнение математических ограничений, устанавливаемых требованиями предметной области, например, равенство энергии Гиббса для сосуществующих фаз (жидкость–газ или жидкость–твердое тело).

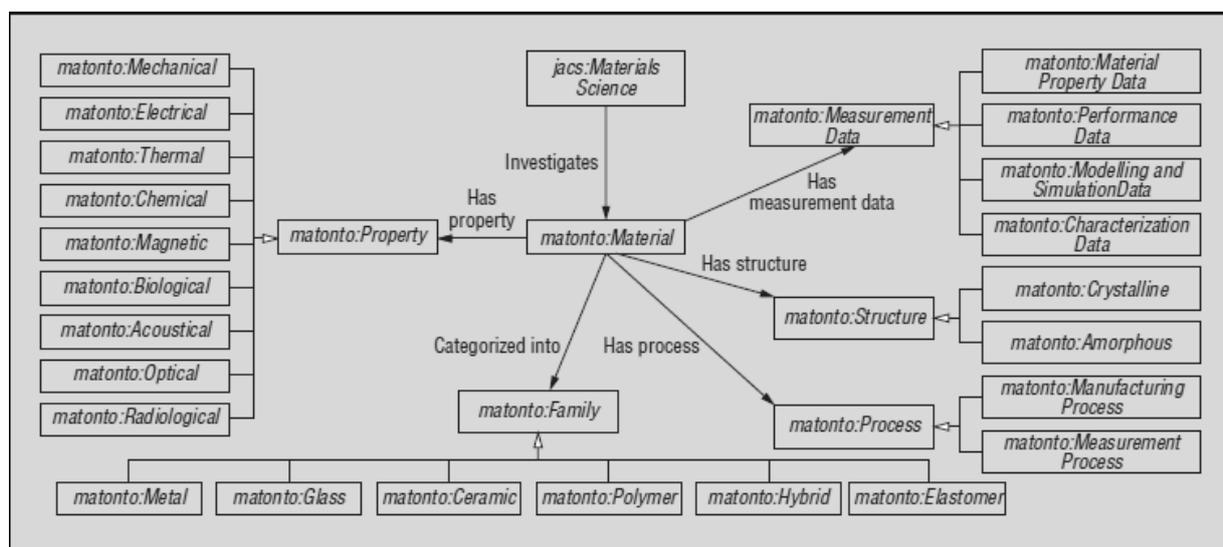


Рис. 8. Схематичное представление верхнего уровня онтологии **MatOnto** [17]

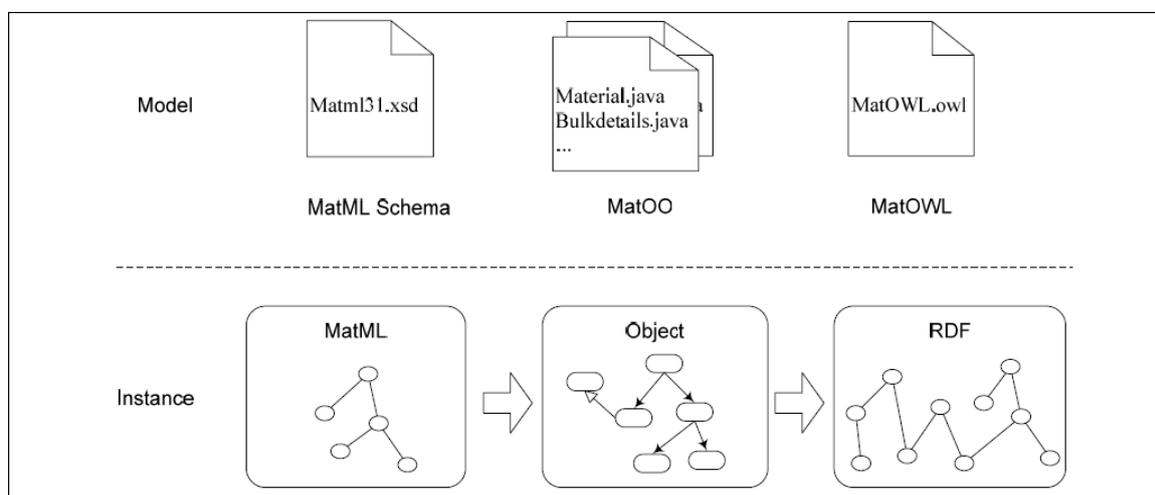


Рис. 9. Схематичное представление конверсии MatML-документа в экземпляры онтологии [18]

Классы онтологии теплофизических свойств веществ [32]

Core classes	
Substances	Определяет вещества, для которых приводятся данные
States	Определяет агрегатные состояния вещества
Properties	Определяет свойства вещества
NumericalData	Определяет набор численных данных для функциональной зависимости свойств вещества в определенных условиях от аргументов
Subsidiary classes	
ConstantsOfSubstance	Определяет набор численных значений констант в определенных условиях среды.
EnvironmentalConditions	Указывает агрегатное состояние вещества и ссылки на набор классов
Uncertainties	Определяет тип погрешности физической величины
Dimensions	Определяет размерности физических величин
Data	Определяет перечень данных из справочников физических величин
DataSource	Определяет источники данных для классов NumericalData , Data
Functions and arguments	
DomainOfFunctionDefinition	Определяет перечень аргументов и ограничений физических свойств для функций
Functions	Определяет функции для вычисления и проверки корректности значений физических величин
ComputingFunc (субкласс Functions)	Определяет перечень функций для вычисления значений свойств
ControlFunc (субкласс Functions)	Определяет перечень проверочных функций, которые контролируют, выполняется ли заданное соотношение при допустимой погрешности

При разработке онтологии активно привлекались внешние источники (существующие онтологии и словари) для унификации семантики. В частности, для именования веществ использован словарь **ChemSpider**, который обеспечивает присвоение веществу уникального идентификатора, например, для водорода **CSID:762** и соответствующего URI www.chemspider.com/Chemical-Structure.762.html. Отдельные термины, связанные с фазами, свойствами, размерностями приняты из онтологий **ChemAxiom** и **QUDT** (*Quantities, Units, Dimensions and Data Types*) [32].

Физические принципы, определяющие свойства веществ, накладывают целую совокупность ограничений на использование понятий. Логические ограничения записаны с использованием конструкций языка OWL. Среди логических ограничений: разбиение класса свойств на два непересекающихся класса (свойства-функции и свойства-константы); обязательность определения аргумента для свойства-функции; согласованность ссылок на состояния вещества с видом свойств-функций (например, запрет на свойство **viscosity** в состоянии **solid**). Математические ограничения задаются отдельно для каждого экземпляра-свойства. Они относятся к свойствам-функциям: требования к области определения, области существования, характеру монотонности и, возможно, другим характеристикам (например, связи двух и более функций). Математические ограничения касаются не

классов, а экземпляров свойств, поскольку области определения и существования функций определяются для каждого свойства отдельно и список свойств допускает расширение. На основании построенной онтологии сгенерирована реляционная БД в СУБД PostgreSQL. В соответствии с 12 классами онтологии создано 12 java-классов, которые отображаются на таблицы реляционной БД данных, рис. 10.

Для системы загрузки данных разработан комплекс, позволяющий анализировать и загружать данные с их проверкой на соответствие онтологии и выполнение ограничений. Задача системы не ограничивается хранением введенных данных, предполагая возможность ряда вычислительных операций, в том числе: для контроля корректности численных данных, основанного на физических принципах; расчета физического свойства в произвольной точке, например, при произвольно заданной пользователем температуры; для случая добавления пользователем нового свойства в виде формулы или программного кода.

Разработанная система заметно облегчила проектирование реляционной БД за счет связи с онтологией. Помимо унификации семантики, онтология обеспечила выполнение логических и математических связей между понятиями и сохранение за пользователем права наращивать списки веществ, свойств, единиц измерения и прочих элементов набора данных при эволюции концептуальной схемы.

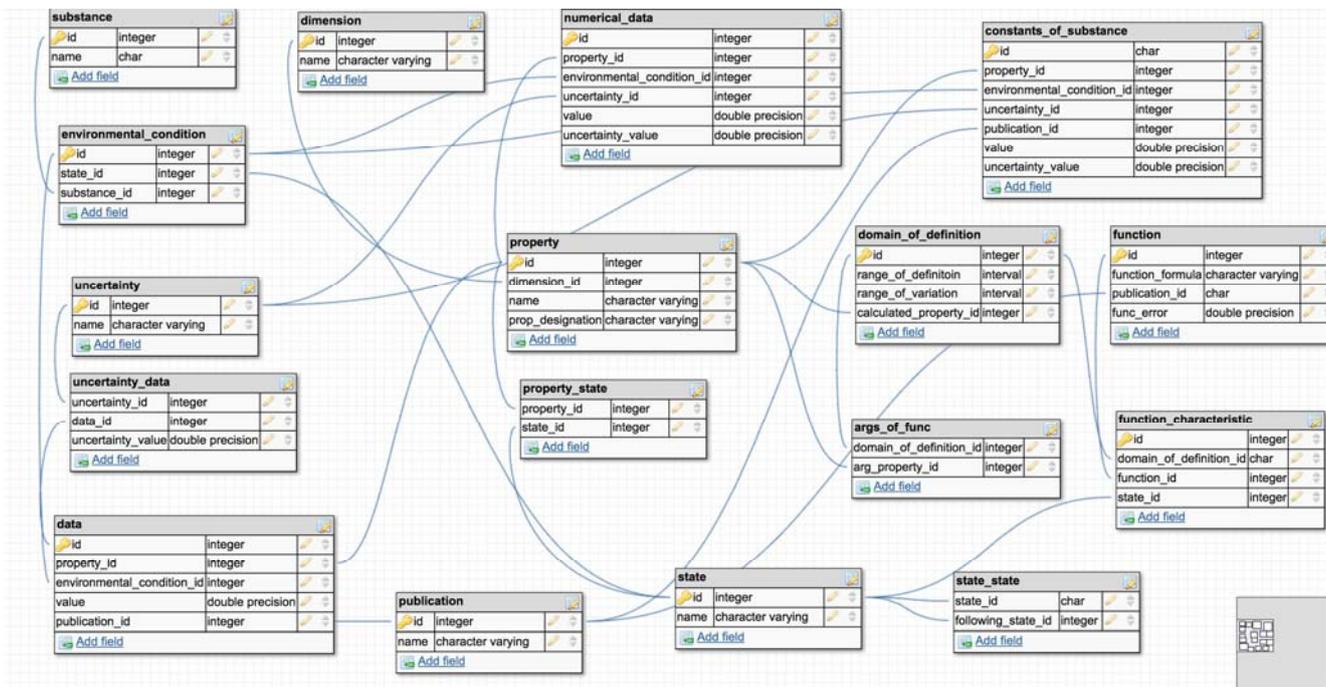


Рис. 10. Таблицы теплофизической БД [32]

ЗАКЛЮЧЕНИЕ

Более десятилетия назад авторы опубликовали статью [2], в которой была обоснована необходимость активного внедрения новых технологий работы с данными для поддержки и эффективного развития общенаучной тематики «Свойства веществ и материалов». Ее специфика, включая многообразие форм представления, требования к совместному хранению «сырых» и обработанных данных, перманентное обновление логической структуры и физических моделей с появлением новых объектов – все это поставило экспертов перед необходимостью отхода от традиционных форм хранения в автономных БД. За прошедшие годы доминирование работы с данными стало тенденцией во многих дисциплинах, называемых теперь *e-science* [4, 5], что, естественно, не обошло и отмеченную нами тематику.

В настоящей работе изучен один из новых подходов к работе с данными по свойствам, использующий различные технологии связывания БД с онтологиями [24–27]. Сами по себе БД, будучи созданы в разных коллективах, с неизбежностью порождают разноразличия в терминологии, логических схемах и форматах данных. Онтология оказалась идеальным средством унификации семантики и подчинения множества концептуальных схем единой структуре, к тому же согласованной в научном сообществе. В результате открывается путь к широкой интеграции структурно или даже тематически разнородных ресурсов с возможностью обмена данными и использования в совместной работе. Нами проанализирован опыт связывания онтологий с БД для ряда областей знания: химии, материаловедения, теплофизики.

В компьютерном сообществе предложено несколько возможных подходов к такому связыванию.

Один из них, названный **DBVO**, предполагает проектирование БД путем отображения онтологии на концептуальную схему БД, т.е. ее согласование с внешним стандартизованным ресурсом. Другая концепция, **OBDB**, имеет прямое отношение к методологии Semantic Web. В этом случае БД используется как хранилище онтологий совместно с экземплярами данных, которые могут быть включены в пространство связанных данных. Наконец, предложена концепция **OBDA** для интеграции разнородных БД за счет использования онтологии как семантического посредника при формулировании запроса между пользователем и БД.

Рассмотренные в работе примеры показывают, что привлечение онтологии, формализующей знания в предметной области, позволяют не только унифицировать и сделать общедоступной семантику предметной области, но и обеспечить распространение знаний, закодированных в логических соотношениях между концептами. Применительно к тематике «Свойства веществ и материалов» формализация знания в онтологии позволяет решить старую проблему [1] – постоянно варьировать форму идентификации новых объектов и номенклатуру свойств и характеристик, что особенно актуально для таких областей, как биоинформатика или нанотехнологии.

СПИСОК ЛИТЕРАТУРЫ

1. Еркимбаев А.О., Зицерман В.Ю., Кобзев Г.А., Фокин Л.Р. Логическая структура физико-химических данных. Проблемы стандартизации и обмена численными данными // Журнал физической химии. – 2008. – Т. 82, № 1. – С. 20–31.
2. Зицерман В.Ю., Кобзев Г.А., Фокин Л.Р. Возможности и перспективы информационных

- технологий в подготовке и распространении справочных данных: свойства веществ и материалов // Научно-техническая информация. Сер. 1. – 2004. – № 2. – С. 7–14.
3. Киселева Н.Н., Дударев В.А., Земсков В.С. Компьютерные информационные ресурсы неорганической химии и материаловедения // Успехи химии. – 2010. – Т. 79, № 2. – С. 162–188.
 4. Hey T. The Fourth Paradigm: Data Intensive Scientific Discovery / eds. T. Hey, S. Tansley, K. Tolle. – Redmond, WA: Microsoft Research, 2009. – 286 p.
 5. Castelli D., Manghi P., Thanos C. A vision towards Scientific Communication Infrastructures. On bridging the realms of Research Digital Libraries and Scientific Data Centers // Int. J. Digit. Libraries. – 2013. – Vol. 13. – P. 155–169.
 6. Еркимбаев А.О., Зицерман В.Ю., Кобзев Г.А., Сон Э.Е., Сотников А.Н. Интеграция баз данных по свойствам вещества. Подходы и технологии // Научно-техническая информация. Сер. 2. – 2012. – № 8. – С. 1–8.
 7. Когаловский М.Р. Энциклопедия технологий баз данных. – М.: Финансы и статистика, 2002. – 798 с.
 8. O'Reily. XML from the inside out. Resource Guide. Science. – URL: www.xml.com/pub/rg/Science.
 9. Murray-Rust P., Rzepa H.S. Markup Languages – How to Structure Chemistry-Related Documents // Chemistry International. – 2002. – Vol. 24, № 4. – P. 9–13.
 10. Елизаров А.М., Липачев Е.К., Малахальцев М.А. Веб-технологии для математика: основы MathML. – М.: Физматлит, 2010. – 192 с.
 11. Frenkel M., Chirico R.D., Diky V. et al. XML-based IUPAC standard for experimental, predicted, and critically evaluated thermodynamic property storage and capture (ThermoML). IUPAC Recommendations 2006 // Pure Appl. Chem. – 2006. – Vol. 78, № 3. – P. 541–612.
 12. Ojala T., Over H.-H. Approaches in using MatML as a common language for materials data exchange // Data Science Journal. – 2008. – Vol. 7. – P. 179–195.
 13. Kaufman J.G., Begley E.F. MatML. A Data Interchange Markup Language // Advanced Materials & Processes/November. – 2003. – P. 35–36.
 14. Park Jun-Hyoung, Song Mi-Young, Yoon Jung-Sik. Design of System for Atomic, Molecular and Plasma-Material Interaction Data Exchange with XSAMS // International Journal of Software Engineering and Its Applications. – 2014. – Vol. 8, № 6. – P. 149–156.
 15. VAMDC-XSAMS reference guide [v 1.0] [r 12.07]. – URL: <http://vamdc.eu/documents/standards/dataModel/vamdcxsams/index.html>.
 16. Ashino T. Materials ontology: an infrastructure for exchanging materials information and knowledge // Data Science Journal. – 2010. – Vol. 9. – P. 54–61.
 17. Cheung Kwok, Hunter Jane, Drennan John. MatSeek: An Ontology-Based Federated Search Interface for Materials Scientists // IEEE INTELLIGENT SYSTEMS. – 2009. – Vol. 24, № 01. – P. 47–56.
 18. Zhang Xiaoming, Hu Changjun, Li Huayu. Semantic Query on materials data based on mapping MatML to an OWL ontology // Data Science Journal. – 2009. – Vol. 8. – P. 1–17.
 19. Degtyarenko K., de Matos P., Ennis M. et al. ChEBI: a database and ontology for chemical entities of biological interest // Nucleic Acids Research. – 2008. – Vol. 36. – Database issue. – D344–D350.
 20. Gruber T.R. A translational approach to portable ontology specification // Knowledge Acquisition. – 1993. – Vol. 5, № 2. – P. 199–220.
 21. Uschold M. Ontologies and Database Schema: What's the Difference? – URL: www.slideshare.net/UscholdM/.
 22. Еркимбаев А.О., Зицерман В.Ю., Кобзев Г.А. Роль метаданных в создании и использовании информационных ресурсов о свойствах веществ и материалов // Научно-техническая информация. Сер. 1. – 2008. – № 11. – С. 13–19.
 23. Frenkel M., Chirico R. D., Diky V. et al. Extension of ThermoML: The IUPAC standard for thermodynamic data communications. (IUPAC Recommendations 2011) // Pure Appl. Chem. – 2011. – Vol. 83, № 10. – P. 1937–1969.
 24. Тузовский А.Ф. Интеграция баз данных на основе онтологий. Слайды симпозиума «Онтологическое моделирование 2010» // Сайт рабочей группы симпозиума «Онтологическое моделирование». – URL: ontology.ipi.ac.ru.
 25. Когаловский М.Р. Системы доступа к данным, основанные на онтологиях // Программирование. – 2012. – № 4. – С. 55–77.
 26. Laallam F.Z., Kherfi M.L., Benslimane S.M. A survey on the complementarity between database and ontologies: principles and research areas // Int. J. Computer Applications in Technology. – 2014. – Vol. 49, № 2. – P. 166–187.
 27. Еркимбаев А.О., Зицерман В.Ю., Кобзев Г.А., Серебряков В.А., Теймуразов К.Б. Технология научных публикаций в среде «открытых связанных данных» // Научно-техническая информация. Сер. 1. – 2013. – № 12. – С. 1–11.
 28. An Y., Hu X., Song I.-Y. Round-Trip Engineering for Maintaining Conceptual-Relational Mappings // Proc. Of 20th International Conference on Advanced Information Systems Engineering (CAiSE'08), LNCS 5074, P. 296–311. Springer, 2008.
 29. Еркимбаев А.О., Зицерман В.Ю., Кобзев Г.А., Серебряков В.А., Шиолашвили Л.Н. Интеграция данных по свойствам веществ и материалов на основе онтологического моделирования предметной области // Журнал «Электронные библиотеки». – 2013. – Т. 16, № 6. – URL: www.elbib.ru/index.phtml?page=elbib/rus/journal/2013/part6/EZKSS.

30. Materials Database Information Portal. – URL: <http://musigny.rds.toyo.ac.jp:8080>.
31. Austin T.S.P., Over H.H. MatDB online – a standards-based system for preserving, managing, and exchanging engineering materials test // Data Science Journal. – 2012. – Vol. 11. – P. ASMD11–ASMD16.
32. Серебряков В.А., Теймуразов К.Б., Хайруллин Р.И., Еркимбаев А.О., Зицерман В.Ю., Кобзев Г.А., Трахтенгерц М.С. Практическая реализация системы интеграции теплофизических данных на основе онтологической модели предметной области // Труды Четвертого Всероссийского симпозиума «Инфраструктура научных информационных ресурсов и систем». (С.-Петербург. 6–8 октября 2014 г.) / под ред. Е.В. Кудашева, В.А. Серебрякова. – М: ВЦ РАН. – Т. 1. – С. 87–111.

Материал поступил в редакцию 18.08.15.

Сведения об авторах

ЕРКИМБАЕВ АДильБЕК ОМИРБЕКОВИЧ – кандидат технических наук, зав. лабораторией баз данных объединенного института высоких температур РАН (ОИВТ РАН), Москва
e-mail: adilbek@ihed.ras.ru

ЗИЦЕРМАН ВЛАДИМИР ЮРЬЕВИЧ – кандидат физико-математических наук, ведущий научный сотрудник лаборатории баз данных ОИВТ РАН
e-mail: vz1941@mail.ru

КОБЗЕВ ГЕОРГИЙ АНАТОЛЬЕВИЧ – доктор физико-математических наук, главный научный сотрудник, советник НИЦ электрофизики и тепловых процессов ОИВТ РАН
e-mail: gkbz@mail.ru

КОСИНОВ Андрей Владимирович – инженер-программист ОКБ ОИВТ РАН
kosinov@gmail.ru

О.М. Нефедов, Л.М. Королева, С.В. Трепалин, Ю.Е. Бессонов, Н.И. Чуракова

Разработка интегрированной системы структурной химической информации*

Рассмотрены молекулярные и реакционные структурные базы данных по химии. Описана история развития программно-технологического комплекса Базы структурных данных по химии ВИНТИ РАН. Показаны новые возможности поиска структурной информации на основе генерации структурных диаграмм.

Ключевые слова: химическая информация, База структурных данных по химии ВИНТИ РАН, поиск химической информации, Интернет-доступ, алгоритмы генерации полициклических структур, структурные диаграммы

ВВЕДЕНИЕ

Неотъемлемой частью химической информации, необходимой специалистам-химикам, является наличие сведений о строении и свойствах веществ. Предметная область химия, наряду с реферативными базами данных, широко представлена в мировом информационном пространстве различными информационными ресурсами, в том числе базами структурных данных по химии, содержащими информацию о структуре и свойствах химических соединений. Крупнейшим центром-генератором химической информации является Chemical Abstracts Service (CAS), отделение Американского химического общества (American Chemical Society). С 1957 года CAS создает специализированную регистрационную систему CAS REGISTRYSM, которая содержит структурную информацию об индивидуальных органических и неорганических химических веществах, о координационных соединениях, о сплавах, минералах, полимерах, солях, а также о смесях химических веществ. Кроме того, в этой информационной системе имеются сведения о более 66 млн биопоследовательностях. Ежедневно в CAS REGISTRYSM регистрируется свыше 15 000 химических веществ [1].

ОСНОВНЫЕ ТЕХНОЛОГИЧЕСКИЕ ЭТАПЫ РАЗВИТИЯ БАЗЫ СТРУКТУРНЫХ ДАННЫХ ПО ХИМИИ ВИНТИ РАН

Последние достижения информационных технологий позволяют значительно совершенствовать и оптимизировать процессы обработки, хранения и поиска химической информации. Так например, фирма

Elsevier создала свой веб-ресурс структурной химической информации - Reaxys[®], который предоставляет доступ к двум базам данных: Beilstein и Gmelin. Первая из них содержит информацию о структуре, свойствах и химических реакциях органических соединений, а вторая – о неорганических соединениях [2]. Краткая характеристика некоторых веб-ресурсов по химии, предоставляющих информацию о молекулярной структуре, химических реакциях и свойствах химических соединений представлены в табл. 1.

В России База структурных данных по химии (далее База СД), являющаяся одной из крупнейших в мире баз данных по химии, генерируется в ВИНТИ РАН. Она содержит более 9 млн химических структур, более 4 млн химических реакций и 15 млн свойств химических соединений. База СД стала формироваться в 1975 г., а в 1998 г. ВИНТИ РАН получил свидетельство (№ 980007 от 26.01.1998г.) об официальной регистрации Базы СД в РосАПО. Ежегодное пополнение Базы СД составляет примерно 180 тыс. соединений и 100 тыс. реакций из более 6 тыс. документов. Основная особенность содержащейся в Базе СД информации заключается в том, что это сведения об индивидуальных низкомолекулярных органических, элементоорганических, координационных соединениях, а также о низкомолекулярных природных соединениях и их синтетических аналогах.

Основные принципы обработки структурной химической информации были разработаны в 1963 г. в ВИНТИ РАН в Отделе «Химия» в виде так называемой фрагментарной произвольно-блочной системы кодирования (или ПБ-код) [3]. ПБ-код – это средство для представления информации о химических соединениях в текстовой форме на основе описания фрагментов химической структуры и взаимосвязей между ними. Такая система кодирования позволяла представить дву- и трехмерные структуры

* Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований, проект № 13-07-00488.

химических соединений в виде линейных записей, которые можно было реализовать на оборудовании и программном обеспечении того времени.

В дальнейшем разработки по созданию автоматизированной информационной системы (АИС) по химии осуществлялись в тесном сотрудничестве с учеными и специалистами ведущих институтов Академии наук, Университетов и вузов СССР, Центрального института информации и документации (ЦИИД) и Народного предприятия «Централизованная обработка информации по химии» (ЦИХ) ГДР. В итоге, к 1972 г. была разработана Автоматизированная система поиска структурной химической информации – ШПРЕЗИ (SPRESI – Speicherung und Recherche strukturischer Information – Хранение и поиск структурной информации). Для функционирования системы ШПРЕЗИ был предложен так называемый фрагментарный код. Формат Шпрези – это бинарный код химического соединения, доступный для чтения только с помощью специального программного обеспечения. При этом химические соединения кодировались вручную с помощью системы произвольно-блочного кода, по которому автоматически строилась таблица связей, а из нее автоматически

вырабатывался фрагментарный код ШПРЕЗИ [4, с. 176–177]. Таким образом, База СД в 1975–1996 гг. создавалась в ВИНИТИ на основе программного обеспечения ШПРЕЗИ.

В 1996 г. в ВИНИТИ РАН была разработана и внедрена новая отечественная программная оболочка CBASE16 (Chemical Base, 16-разрядная версия), предназначенная для ввода и обработки структурной химической информации на основе графического ввода структурных данных с использованием структурного редактора CHED, разработанного С.В. Трепалиным [5–7]. Программная оболочка CBASE16 позволила профессионально и качественно осуществлять ввод, обработку, хранение и поиск информации о химических соединениях и реакциях, в которых эти соединения принимают участие. Однако бурное развитие информационных технологий, программного обеспечения и вычислительной техники привели к тому, что разработанная программная оболочка CBASE16 морально устарела. Поэтому в ВИНИТИ РАН был разработан программный комплекс CBASE32, который представляет собой усовершенствованную 32-разрядную версию программной оболочки CBASE.

Таблица 1

Молекулярные и реакционные структурные базы данных по химии (данные 2015 г.)

Название	Кол-во соединений	Кол-во реакций	Производитель	Адрес веб-ресурса
CAS REGISTRY SM CASREACT [®]	101 млн	67,5 млн	CAS*	https://www.cas.org
Beilstein	11,7 млн	23 млн	Elsevir	http://www.elsevier.com/solutions/reaxys
Gmelin	2,4 млн	1,8 млн	Elsevir	http://www.elsevier.com/solutions/reaxys
Spresi	5,7 млн	4,4 млн	Infochem	http://www.infochem.de/products/database/s/spresi.shtml
Index Chemicus [®]	2,6 млн		ISI Thomson Scientific	http://thomsonreuters.com/en/products-services/scholarly-scientific-research/scholarly-search-and-discovery/index-chemicus.html
Current Chemical Reactions [®]		880 тыс.		http://thomsonreuters.com/en/products-services/scholarly-scientific-research/scholarly-search-and-discovery/current-chemical-reactions.html
Cambridge Structural Database	750 тыс.		The Cambridge Crystallographic Data Centre	http://www.ccdc.cam.ac.uk
Inorganic Crystal Structure Database	177 тыс.		FIZ-Karlsruhe	https://icsd.fiz-karlsruhe.de

* В CAS REGISTRY в 2014 г. было зарегистрировано больше химических веществ, чем за весь период с 1965 по 1990 гг.

Существенным отличием CBASE32 от предыдущей версии является представление и автоматический ввод в базу данных не только одностадийных, но и многостадийных химических реакций. Такое представление реакций позволяет максимально полно описывать экспериментальные детали из оригинального документа и отражать реальную последовательность проведения стадий реакции, приводящих к получению целевого продукта [8]. Кроме этого, для адекватного и однозначного представления стереохимической информации была усовершенствована система автоматического распознавания конфигурации асимметрических центров химических соединений, учитывающая особенности пространственного строения практически всех типов индексируемых соединений [9]. Разработанное программное обеспечение позволило также реализовать уникальные информационные модели координационных соединений и элементоорганических соединений, которые значительно превосходят по своим описательным возможностям принятые в мировой практике стандарты. Для повышения качества формирования Базы СД и эффективности поиска структурной химической информации программный комплекс CBASE32 был дополнен новым компонентом – справочником химических соединений, наиболее часто встречающихся в химической литературе [10]. Усовершенствованный программно-технологический комплекс обработки, хранения, распознавания, использования химической информации при создании Базы СД CBASE32 позволяет эффективно использовать программы поиска, как точных химических структур, так и структурной химической информации по фрагменту структуры. Разработанные и внедренные средства поиска структурной химической информации сделали возможным осуществить интерактивный доступ через Интернет к имеющемуся в Базе СД массиву данных [11, 12]. Относительные объемы структурной химической информации и форматы их представления в Базе СД по химии ВИНТИ РАН приведены в табл. 2.

Как видно из табл. 2, более 60% химических структур, содержащихся в Базе СД, было обработано до середины 90-х гг. В конце 90-х гг. были разработаны конверторы для представления структур, хра-

нящихся в форматах ПБ-кода и Шпрези, в виде Т-графов. Т-графом называется линейная (текстовая запись) структуры химического соединения (молекулярного графа) [13]. Т-граф позволяет кодировать как органические, так и неорганические и комплексные соединения. Большинство структур ретрофонда (накопленных за период с 1975 г. по 1995 г.) в результате проведенных работ хранятся в виде Т-графов.

Известно, что отличительной особенностью химической информации является одновременно не только ее быстрый рост, но и медленное старение. При этом главные объекты исследований – химические соединения и реакции, а практические задачи – это синтез новых соединений с необходимыми или заданными свойствами и разработка новых химических реакций на основе уже известных. Определение современного состояния научных исследований и разработок в области химии не зависит от глубины поиска специальной и профессиональной информации.

Данные в форматах CBASE16 и CBASE32 содержат информацию о рисунках молекулярных структур и могут выгружаться в файлы форматов SDF (для структур) и RDF (для реакций). В настоящее время разработан конвертор для преобразования информации из формата CBASE16 в формат CBASE32, и данные, накопленные в период с 1996 г. по 2009 г., переводятся в формат CBASE32 с использованием этого конвертора. Отсутствие информации об изображениях структурных формул в ретрофонде за период 1975 – 1995 гг. требует преобразования в единый формат CBASE32.

Для полноценного представления химического соединения необходимо иметь графическое изображение структурной формулы. Имеется возможность конвертировать данные об атомах и топологии молекул, представленных ПБ-кодом или Т-графами, в формат CBASE32 или SDF. Однако автоматическое формирование приемлемого двумерного изображения (визуализация) структурной формулы молекулы только на основании топологии (таблицы связей) является достаточно сложной задачей. При этом, визуализация структур ретрофонда должна преобразовать максимально возможное число структур с минимальными затратами и потерями.

Таблица 2

**Общая характеристика структурной химической информации
Базы СД ВИНТИ РАН**

Период	Структуры химических соединений (%)	Химические реакции (%)	Форма представления данных
1975г.-1995г.	62,8	70,7	ПБ-код, Шпрези, Т-граф
1996г.-2009г.	27,0	20,1	CBASE16, Т-граф
2010г.-2014г.	10,2	9,2	CBASE32, SDF, RDF
Итого	100	100	

ПРОБЛЕМЫ ВИЗУАЛИЗАЦИИ: БИБЛИОТЕЧНЫЙ ПОДХОД

Библиотечный подход известен достаточно давно [14]. Суть его состоит в том, что создается набор (библиотека) заранее прорисованных шаблонов. Обычно шаблоны соответствуют циклическим фрагментам химической структуры, и их рисунки удовлетворяют общепринятым требованиям к изображениям химических структур. Топология молекулы анализируется на предмет наличия фрагментов, изоморфных шаблонам из библиотеки. После того как все фрагменты идентифицированы, происходит сборка изображения всей структуры [15, 16]. На этом принципе был основан экспериментальный программный комплекс, разработанный в ВИНТИ РАН в 1993 г. Здесь в качестве шаблонов были использованы так называемые химические блоки, представляющие собой максимальные связанные циклические компоненты молекулярного графа. Комплекс позволял достаточно качественно автоматически рисовать структуры химических соединений, если все химические блоки входили в состав библиотеки шаблонов. Если какой-то химический блок в библиотеке отсутствовал, то он автоматически рисовался при помощи эвристического алгоритма. Но, как правило, такой рисунок требовал последующей ручной коррекции. Исследования, проведенные на массиве данных 1992 г. по визуализации химических структур с помощью этого экспериментального программного комплекса, показали, что число различных химических блоков стабильно растёт с увеличением количества соединений в базе данных (рис. 1). На оси абсцисс рис. 1 представлены номера РЖ Химия за 1992 г., соответ-

ствующие порядку добавления структур в Базу СД в течение года. Каждый номер представляет около 20000 структур. На оси ординат – суммарное число уникальных химических блоков, необходимых для визуализации введенных в Базу СД структур химических соединений. Экстраполируя кривую на всю базу данных (9 161 021 соединений) получаем величину, приблизительно равную 250 000 химических блоков¹.

Рассмотрим подробнее разработанный нами метод построения изображений полициклических соединений, также основанный на библиотечном принципе. Однако библиотека шаблонов такого метода имеет существенно меньший размер (порядка 200 фрагментов).

ГЕНЕРАЦИЯ СТРУКТУРНЫХ ДИАГРАММ В ПОЛИЦИКЛИЧЕСКИХ СОЕДИНЕНИЯХ

Проведенные научно-исследовательские работы по переводу информации 1975 – 1995 гг. в формат CBASE32 позволили разработать и апробировать механизм генерации структурных диаграмм в полициклических соединениях.

Для демонстрации структур, хранимых в виде линейных кодов, необходимо генерировать 2D-декартовы координаты атомов – структурную диаграмму. Очевидно, что эта задача имеет бесконечное множество решений. При генерации структурных диаграмм основной критерий – вид рисунка структуры. Химические структуры выглядят наиболее наглядно, если все длины связей эквивалентны, а углы между связями равны $2\pi/3$. Это не всегда можно выполнить, но к этому надо стремиться. В литературе описаны алгоритмы генерации структурных диаграмм [17, 18].

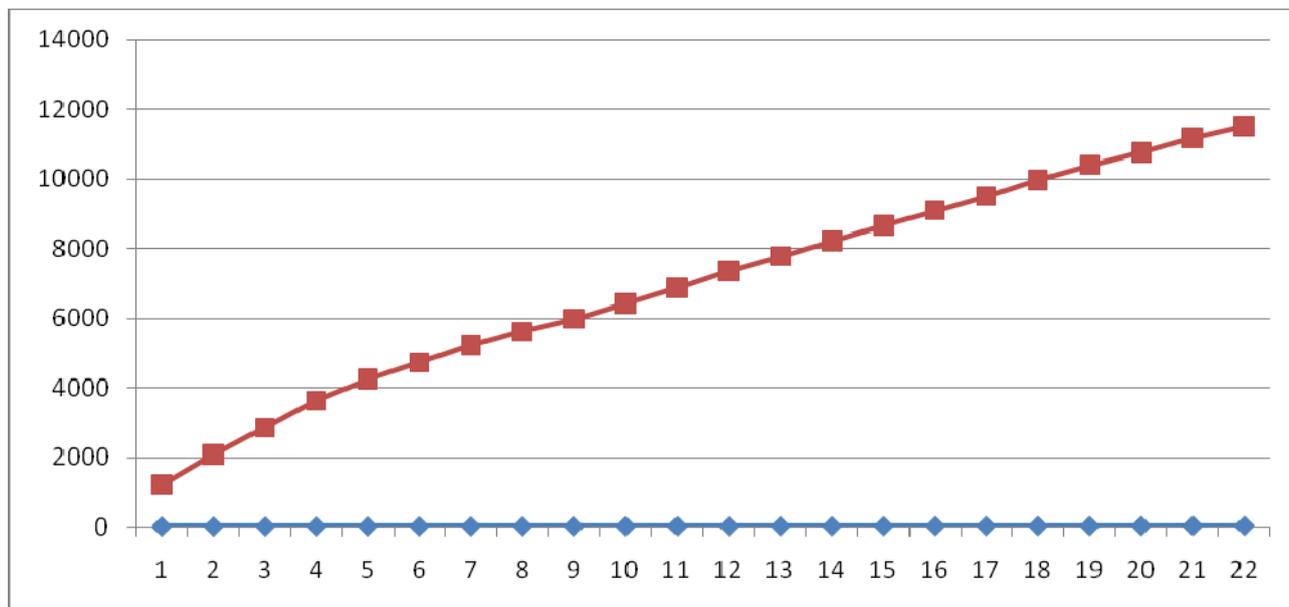


Рис. 1. Количество уникальных химических блоков-шаблонов на номер РЖ Химия за 1992 г.

¹ Это верно, если нет «насыщения», т.е. число уникальных химических блоков не становится в какой-то момент константой. Изучение данного вопроса представляет научный интерес и является отдельной темой.

Большинство редакторов химических структур имеют команду “Clean Structure” [19–21]. При выполнении этой команды генерируется структурная диаграмма, в которой чаще всего одинаковые длины связей и оптимальные валентные углы. Однако в трех коммерческих наиболее часто используемых редакторах химических структур – ISIS/Draw [20], ChemDraw [21], ChemSketch [19] и свободно распространяемой JChemPaint [22] возникают проблемы при генерации полициклических структур. На рис. 2 показан кубан (C_8H_8), в котором 2D-координаты атомов определены случайным образом. Для такой “возмущенной” структуры командой “Clean Structure” генерировалась оптимальная структура в разных химических редакторах. Полученные результаты показаны на том же рисунке. Как видно, только ChemSketch (ACD Labs) генерирует подходящие координаты атомов. Однако и ChemSketch не может создать оптимальной структуры при усложнении задачи.

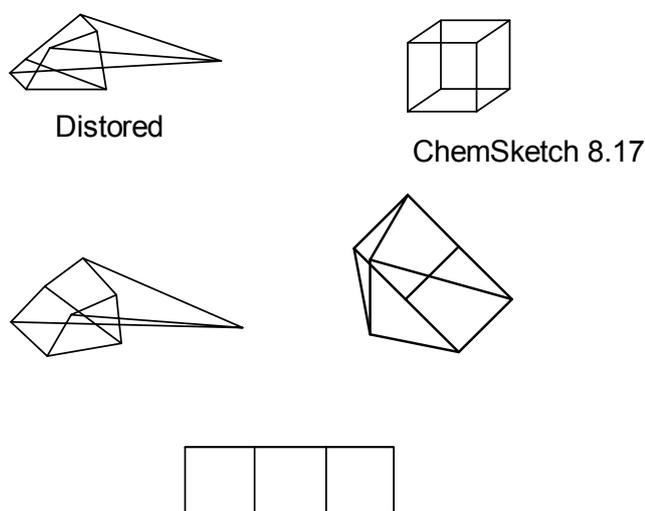


Рис. 2. Результат выполнения Clean команды в популярных программах— ChemSketch 8.17, ISIS/DRAW 2.5, ChemDraw Ultra 8.0, JChemPaint 2.0.6 Pre.

Для решения проблемы показа полициклических структур предлагается использовать небольшую библиотеку фрагментов с заранее заданными координатами атомов. Фрагменты из библиотеки ищутся на вложение в соединение, и, если удастся найти, – то всем найденным атомам присваиваются координаты атомов во фрагменте с подходящим масштабирующим фактором, сдвигом и поворотом для достижения оптимальной конфигурации. Поскольку фрагменты используются для разрисовки любых структур, то все атомы и связи считаются эквивалентными и совместимыми с любыми атомами (связями) в структуре. Соответственно, можно создать компактное хранилище данных и это является важным для генерации 2D-координат.

Ниже приведены примеры используемых фрагментов (рис. 3). Фрагменты включают в себя как полициклические структуры (адамantan, норадамantan), так и большие циклы (циклооктадекан). Включение больших циклов делает их разрисовку более наглядной. При рисовании цикла как выпуклого полиэдра размером более 12 атомов перестают быть заметными углы меж-

ду связями на атомах углерода, и визуальная оценка размера цикла из картинки становится затруднительной. Если имеется несколько способов вложения фрагмента в структуру, то соответствие атомов (mapping) после поиска вложения графа является случайным. Любые симметричные фрагменты имеют несколько способов вложения фрагмента в структуру. При генерации структурных диаграмм к заданному фрагменту могут присоединяться другие фрагменты.

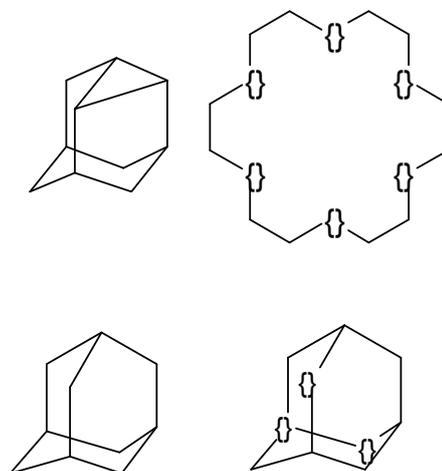


Рис. 3. Примеры фрагментов используемых для генерации 2D-координат атомов. Атомы, которые не могут быть связаны с другими атомами, имеют метку {}.

При этом местом их присоединения могут быть выбраны произвольные места в эквивалентных атомах, в том числе и “густые” – где содержится много атомов и связей, и добавлять новые не представляется возможным. Примеры плохо сгенерированных структурных диаграмм приведены на рис. 4. Для решения этой проблемы во фрагментах помечаются атомы, к которым нельзя присоединять другие фрагменты. Такой фрагмент показан на рис. 3 – адамantan с тремя помеченными атомами и циклооктадекан. Но и сам адамantan сохраняется в библиотеке фрагментов для возможной генерации структурной диаграммы полизамещенного адаманта.

При поиске вложения фрагмента в структуру, помеченные атомы рассматриваются как точно координированные. Найденные в химической структуре атомы должны иметь то же самое количество соседей и присоединенных связей, как это указано во фрагменте. При таком поиске гарантируется, что к “густым” атомам при генерации структурных диаграмм не будут добавлены новые соседи. Для корректного поиска библиотека должна быть отсортирована по размеру фрагментов. Поиск начинается с фрагмента, содержащего максимальное количество атомов, а при равенстве количества атомов, нужно начать с фрагмента с максимальным количеством связей. Если имеются как фрагмент с атомами, помеченными для отсутствия присоединения других фрагментов, так и без помеченных атомов, поиск начинается с помеченного фрагмента. В противном случае, можно удачно разрисовать небольшие участки соединения и не найти максимально близкого к структуре фрагмента.

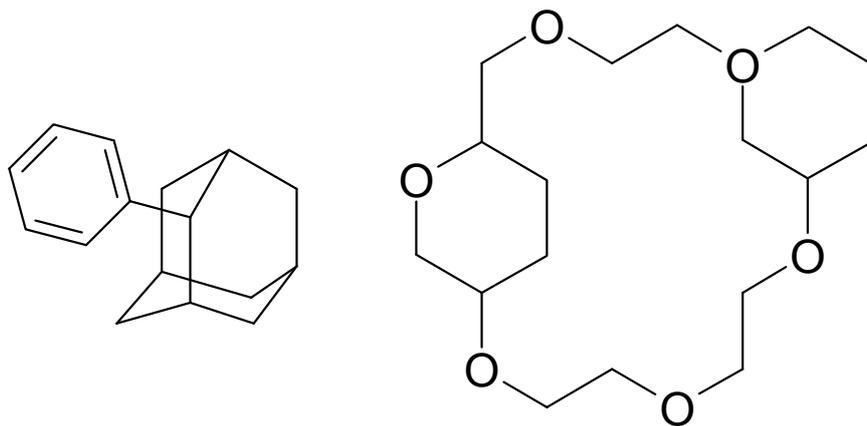


Рис. 4. Пример некорректно сгенерированных 2D-атомных координат

Следующая проблема, которую следует решить - это использование нескольких фрагментов из библиотеки для генерации структурных диаграмм или многократное использование одного и того же фрагмента. Для ее решения требуется огромная библиотека фрагментов. Например, чтобы нарисовать 1,1'-диадамантан потребуется его хранение целиком в библиотеке фрагментов. Для этого следует использовать многократный поиск фрагментов в целевой молекуле. Соответственно, если координаты атомов уже восстановлены, то они отмечаются и становятся недоступными для поиска. Отмеченные атомы дают "false" при попытке совместить их с любым из атомов фрагмента в базе.

Последовательность описания структуры по фрагментам можно представить в виде следующих этапов.

1. Вычисляется набор минимальных циклов в соединении, используя алгоритм из [23], и сохраняется в списке.

2. Осуществляется поиск по библиотеке фрагментов и при нахождении фрагмента восстанавливаются координаты найденных атомов. Если фрагмент не найден, то рисуется цикл максимального размера из списка, полученного в пункте 1. Если циклы отсутствуют - то пара связанных максимально координированных атомов используется как первоначальный фрагмент для генерации структурных диаграмм.

3. Реализуется поиск фрагмента в библиотеке среди атомов, координаты которых не определены. Этот фрагмент должен быть соединен связью с атомом, 2D-координаты которого уже известны.

4. Если фрагмент найден, то координаты атомов структуры, которые соответствуют найденному фрагменту, становятся известными. Далее переход к пункту 3 - ищется следующий фрагмент. Если ни одного фрагмента, который связан связью с атомом с известными координатами, не найдено - переход к пункту 5.

5. Все циклы из списка (пункт 1), для которых имеется хотя бы один атом с известными координатами, добавляются к структуре.

Если известны координаты одного атома, то добавляется спироцикл со стандартными длинами связей и углами, определяемыми из его размера. Если известны координаты двух атомов, соединенных связью, то добавляется конденсированный цикл с длиной связи, равной известной длине связи и углами $2\pi/N$ (N - размер цикла). Если известны координаты трех и более атомов (полицикл), то замыкается цепочка, при этом, по возможности, новые атомы располагают таким образом, чтобы связи не пересекались с уже имеющимися связями. Далее - переход к пункту 3. Если координаты ни одного атома не были определены в этом пункте - переход к пункту 6.

6. Вычисляются координаты атомов, связанных ациклическими связями с атомами, координаты которых известны. Используется стандартная длина связи и гексагональная система координат (угол между осями $2\pi/3$). Если имеется длинная цепочка атомов - то вычисляются координаты только одного атома, связанного с атомом, координаты которых были вычислены при выполнении пунктов 2-5. Если координаты хотя бы одного атома определяются в этом пункте, то переход к пункту 3, в противном случае - к пункту 7.

7. Проверяется, были ли определены координаты всех атомов. Если да - то процесс завершается. Если нет - то соединение - это несвязанный граф, содержащий два или несколько фрагментов. Процесс повторяется, начиная с пункта 1, для следующего фрагмента - и так до их полного исчерпания.

Сгенерированные структурные диаграммы для некоторых полициклических соединений приведены на рис. 5.

Безусловно, можно найти полициклические структуры, для которых описанный здесь алгоритм генерирует плохие диаграммы. Однако их можно добавить в библиотеку фрагментов и далее генерировать качественные структурные диаграммы.

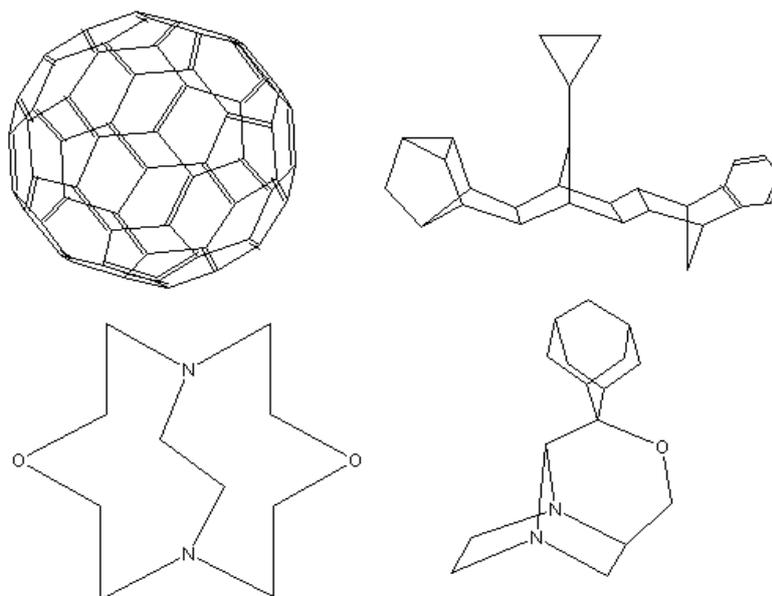


Рис. 5. Пример структурных диаграмм, сгенерированных из линейных строк

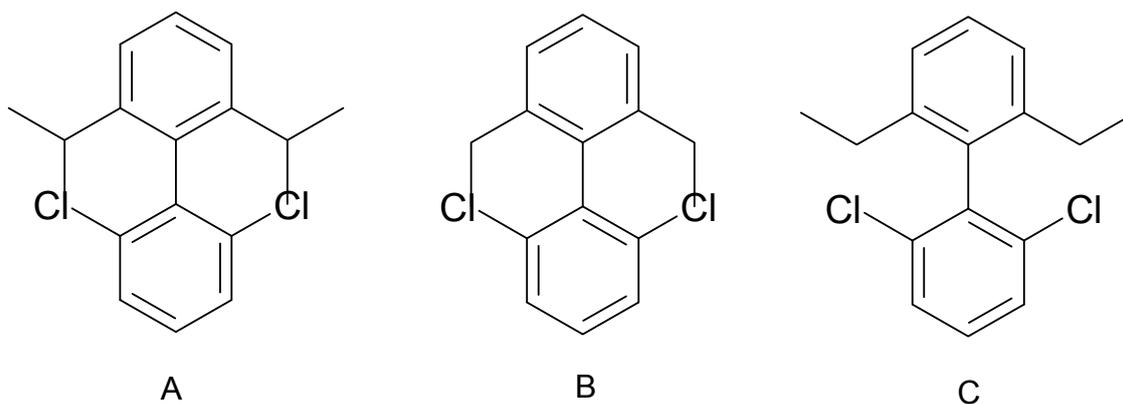


Рис. 6. Структуры, содержащие перекрывающиеся фрагменты (А, В), и структура без перекрывания (С).

ПЕРЕКРЫВАЮЩИЕСЯ ФРАГМЕНТЫ

При генерации структурных диаграмм в случае, если содержатся объемные структурные заместители, то часто отдельные атомы и связи перекрываются. Во многих случаях перекрывания можно избежать, если некоторые фрагменты вращать вокруг ациклических связей на 180 градусов. Однако, например, для 2,6-диизопропил-2',6'-дихлорбифенила, как показано на рис. 6 А, перекрывания атомов хлора с метильной группой не удастся избежать при помощи любых поворотов. В этом случае приходится изменять длины некоторых связей или использовать нестандартные углы между связями, что приводит к генерации структурной диаграммы низкого качества. В то время как для перекрывающегося 2,6-диэтил-2',6'-дихлорбифенила (рис. 6 В) посредством вращений вокруг связей С-С можно сгенерировать структурную диаграмму высокого качества (рис. 6 С), в которой отсутствуют перекрывающиеся атомы и связи.

Для решения этой проблемы использовался слегка упрощенный алгоритм – а именно, вращение вокруг ациклических связей. Первоначально при генерации структурной диаграммы новые атомы цепочки добавляются в направлении ее роста. При этом получаются цепочки максимальной длины. Если при этом имеются перекрывающиеся фрагменты, то есть шанс убрать перекрывание подходящим вращением вокруг ациклических связей. Первоначально для всех ациклических связей определяется количество неэквивалентных заместителей. В этом случае для каждого атома в молекуле берется сферический фрагмент, как это описано для генерации HOSE кода [24], причем радиус этого фрагмента берется равным топологической длине молекулы. После этого считаются атом-центрированные индексы [25]. Если индексы совпадают для пары атомов, присоединенных к одному из атомов при ациклической связи – то они являются эквивалентными и вращение вокруг такой связи бесполезно. Далее формируется список таких связей, и

осуществляется вращение на 180 градусов для всех возможных комбинаций вращений. Общее число вращений равно 2^N (N – число вращающихся связей). Чтобы время вычислений было не очень долгим, максимальное число вращающихся связей ограничили 16. В рамках данного исследования не проводилась коррекция структурной диаграммы с перекрытыми связями и атомами для фрагментов с большим числом вращающихся связей.

ЗАКЛЮЧЕНИЕ

1. Внедрение новых информационных технологий обработки структурных химических данных, разработка современного программно-технологического комплекса обеспечивают рациональное использование всего информационного массива Базы СД ВИНТИ РАН за более чем 40-летний период.

2. Специфическая особенность Базы СД ВИНТИ РАН заключается в строгом соблюдении принципов отбора, обработки и хранения информации об индивидуальных низкомолекулярных органических, элементоорганических, координационных соединениях, а также о низкомолекулярных природных соединениях и их синтетических аналогах.

3. Решена задача генерации структурных диаграмм в полициклических соединениях из линейных кодов в графическое изображение структуры химического соединения.

4. В рамках совершенствования лингвистического обеспечения Базы СД для стандартизации и нормализации описания структурных химических данных сформирован глоссарий, имеющий самостоятельное значение для специалистов не только как справочник, но и для организации формально-логического контроля на всех стадиях обработки информации.

5. Проведенная модернизация Базы СД ВИНТИ РАН позволяет привести имеющийся массив информации к единообразию и обеспечить возможность его интерактивного использования для разнохарактерного контингента пользователей.

СПИСОК ЛИТЕРАТУРЫ

1. Chemical Substances – CAS REGISTRY. – URL: <https://www.cas.org/content/chemical-substances> (дата обращения: 14.09.2015)
2. Reaxys and Reaxys Medical Chemistry. Chemistry research solutions. – URL: <http://www.elsevier.com/solutions/reaxys> (дата обращения: 14.09.2015)
3. Влэдуц Г.Э., Гейвандов Э.А. Автоматизированные информационные системы для химии. – М.: Наука, 1974. – 310 с.
4. Черный А.И. Всероссийский институт научной и технической информации: 50 лет служения науке. – М.: ВИНТИ, 2005. – 316 с.
5. Алфимов М.В., Авакян В.Г., Трепалин С.В., Воронезева Н.И., Чуракова Н. И. Универсальная программная оболочка для создания баз данных химических соединений и реакций // Докл. РАН. – 1999. – Т. 366, № 5. – С. 639-642.
6. Trepalin S.V., Yarkov A.V. CheD – chemical database compilation tool, Internet server and client for SQL servers // J. Chem. Inf. Comput. Sci. – 2001. – Vol.41. – P.100-107.
7. Trepalin S.V., Gerasimenko V.A., Kozyukov A.V., Savchuk N.Ph., Ivaschenko A.A. New Diversity Calculations Algorithms Used for Compound Selection // J. Chem. Inf. Comput. Sci. – 2002. – Vol. 42. – P.249-258.
8. Воронезева Н.И., Трепалин С.В., Чуракова Н.И., Нечаева К.С., Королева Л.М. Система представления и ввода информации о многостадийных химических реакциях с помощью программного комплекса CBASE32 // Научно-техническая информация. Сер. 2. – 2005. № 7. – С. 7 – 11.
9. Немировская И.Б., Трепалин С.В., Королева Л.М. Представление стереохимической информации в Базе структурных данных ВИНТИ РАН // Научно-техническая информация. Сер. 2. – 2006. – № 4. – С. 1 – 6.
10. Воронезева Н.И., Трепалин С.В., Чуракова Н.И. Нечаева К.С., Королева Л.М. Глоссарий как элемент стандартизации ввода данных в программном комплексе CBASE32. // Научно-техническая информация. Сер. 2. – 2007. – № 6. – С. 19 – 24.
11. Нефедов О.М., Трепалин С.В., Королева Л.М., Бессонов Ю.Е. Быстрый поиск точных химических структур в больших базах данных с использованием InChI Key кодировки структур // Научно-техническая информация. Сер. 2. – 2013. – № 12. – С. 27 – 33.
12. Нефедов О.М., Трепалин С.В., Королева Л.М., Бессонов Ю.Е., Чуракова Н.И. База структурных данных по химии ВИНТИ РАН: проблемы поиска по фрагменту структуры // Научно-техническая информация. Сер. 2. – 2014. – № 12. – С. 19 – 29.
13. Арефьев В.Б. О некоторых особенностях математической концепции представления молекулярных графов в регистрационной системе химических соединений ВИНТИ // В сб. "Материалы конференции НТИ-99". – М.: ВИНТИ, 1999. – С.24-26.
14. Carhart R.E. A model-based approach to the teletype printing of chemical structures // J. Chem. Inf. Comput. Sci. – 1976. – Vol.16. – P.82-88.
15. Dittmar P.C., Moskus J., Couvreur K.M. An algorithmic computer diagrams // J. Chem. Inf. Comput. Sci. – 1977. – Vol.17, № 3. – P.186 – 192.
16. Леоненко Ю.П. Программный комплекс для автоматического изображения молекулярных структур // В кн.: Вопросы алгоритмического анализа структурной информации. Вычислительные системы. Т. 119. – Новосибирск, 1987. – С. 102-111.
17. Nelson H.E. Structure Diagram Generation. // Reviews in Computational Chemistry. – 1999. – Vol. 13. – P. 313-398.

18. Fricker P.C., Gastreich M., Rarey M. Automated Drawing of Structural Molecular Formulas under Constraints // J. Chem. Inf. Comput. Sci. – 2004. – Vol. 44. – P. 1065-1078.
19. MDL/CrossFire. – URL: <http://www.mimas.ac.uk/crossfire/autonom.html> (дата обращения: 24.08.2015)
20. MDL™ ISIS Draw 2.5. – URL: http://www.mdl.com/products/framework/isis_draw/index.jsp (дата обращения: 24.08.2015)
21. Cambridge ChemDraw Ultra 8.0. – URL: <http://www.cambridgesoft.com/products/family.cfm?FID=2> (дата обращения: 24.08.2015)
22. Steinbeck C., Han Y., Kuhn S., Horlacher O., Luttmann E., Willighagen E. The Chemistry Development Kit (CDK): An Open-Source Java Library for Chemo- and Bioinformatics // J. Chem. Inf. Comput. Sci. – 2003. – Vol. 43. – P. 493-500.
23. Figueras J. Ring Perception Using Breadth-First Search // J. Chem. Inf. Comput. Sci. – 1996. – Vol. 36. – P.986-991.
24. Bremser W. HOSE - a novel substructure code // Anal. Chim. Acta. – 1978. – Vol. 103. –P. 355-365.
25. Trepalin S.V., Yarkov A.V., Dolmatova L.M., Zefirov N.S., Finch S.A.E. WinDat: An NMR Database Compilation Tool, User Interface and Spectrum Libraries for Personal Computers // J. Chem. Inf. Comput. Sci. – 1995. – Vol. 35. – P. 405-411.

Материал поступил в редакцию 16.09.15.

Сведения об авторах

НЕФЕДОВ Олег Матвеевич – академик РАН, доктор химических наук, советник Президиума РАН, Москва
e-mail: onefedov@ras.ru

КОРОЛЕВА Любовь Михайловна – кандидат химических наук, зав. отделением научной информации по проблемам химии и наук о материалах ВИНИТИ РАН, Москва
e-mail: lkorol@viniti.ru

ТРЕПАЛИН Сергей Владимирович – кандидат химических наук, ведущий научный сотрудник ФГБУН Институт физиологически активных веществ РАН
e-mail: trep@chemical-block.com

БЕССОНОВ Юрий Ефимович – кандидат технических наук, зав. Отделом программного обеспечения и сопровождения информационных систем по химии ВИНИТИ РАН
e-mail: bessonov-ye@rambler.ru

ЧУРАКОВА Наталия Исааковна – кандидат химических наук, зав. сектором структурной информации по биоорганической химии ВИНИТИ РАН
e-mail: nichurak@rambler.ru

Модель порождения орграфов информации по орграфу метаинформации для двухуровневой модели сложноструктурированных информационных единиц*

Для орграфовой связной двухуровневой модели информационных единиц описывается модель порождения орграфов информации по орграфу метаинформации, которая может быть положена в основу компьютерных средств редактирования инвариантной по отношению к технологическим пространствам (предметным областям) сложноструктурированной информации различных уровней абстракции в привычной для ее носителей системе понятий – без специального обучения последних или участия в этом процессе специалистов-посредников.

Ключевые слова: *двухуровневая орграфическая модель представления информации, порождение орграфов, управляемое метаинформацией, поверхностное клонирование орграфов, глубокое клонирование орграфов, зависимое клонирование орграфов*

ВВЕДЕНИЕ

Орграфическая связная двухуровневая модель информационных единиц, соответствующая метафоре анкетирования, описана в работе [1]. Для этой модели предложены языки представления орграфов информации и метаинформации; модель направлена на преодоление трудностей, с которыми сталкиваются пользователи систем извлечения экспертных знаний [2–8], в основе которых лежит объектно-ориентированная парадигма представления метаинформации и информации [9–12].

В настоящей работе описываются модель соответствия между орграфами информации и метаинформации, модель порождения орграфов информации по орграфу метаинформации, а также методы клонирования вершин и подграфов орграфов информации. Рассматриваются два способа формирования орграфов, представляющих сложноструктурированную информацию, – путем их порождения по орграфу, представляющему метаинформацию, а также путем зависимого и независимого клонирования подграфов орграфов, представляющих информацию. При этом в обоих случаях обеспечивается соответствие между орграфами информации и метаинформации.

СООТВЕТСТВИЕ МЕЖДУ ОРГРАФАМИ ИНФОРМАЦИИ И МЕТАИНФОРМАЦИИ

В орграфической связной двухуровневой модели анкеты [1] начальной вершине орграфа информации соответствует начальная вершина орграфа метаинформации.

Орграф информации соответствует орграфу метаинформации (т.е. представляет систему экземпляров понятий для системы понятий, представленной орграфом метаинформации, или, что то же самое, представляет формально правильно заполненную анкету), если существует такое однозначное всюду определенное соответствие между множеством дуг орграфа информации и множеством дуг орграфа метаинформации, что выполнены следующие условия:

- если из вершины v_0 орграфа информации, которой соответствует вершина w_0 орграфа метаинформации, выходит дуга d , то ей соответствует одна из дуг D , у которой спецификатор имеет значение, отличное от «пропуск», и которая выходит из вершины w_0 орграфа метаинформации; при этом конец дуги D соответствует концу дуги d ;
- если из вершины метаинформации w_0 , которая соответствует вершине v_0 орграфа информации, выходит путь $w_0 \Rightarrow^{D_1} w_1 \Rightarrow^{D_2} \dots \Rightarrow^{D_{n-1}} w_{n-1} \Rightarrow^{D_n} w_n$, в котором у дуг D_1, \dots, D_{n-1} спецификатор имеет значение «пропуск», а у дуги D_n спецификатор имеет значение, отличное от «пропуск», из вершины v_0 вы-

* Работа выполнена при частичной поддержке РФФИ (проект 14-07-00299 и проект 15-07-03193) и частичной поддержке программы «Дальний Восток»

ходит дуга $d_i (v_0 \Rightarrow^d_i v_i)$, то дуге d_i соответствует одна из дуг D_n , входящая в вершину w_n , а w_n соответствует вершине v_i ;

- если вершине v орграфа информации соответствует вершина w орграфа метаинформации и вершина w имеет поля «имя» со значением n и «сорт» со значением s , то вершина v имеет поле «значение», значением которого является произвольная константа c^* сорта s ;

- если вершине v орграфа информации соответствует вершина w орграфа метаинформации и вершина w имеет поле «значение» со значением c , то вершина v имеет поле «значение» со значением c ;

- если вершине v орграфа информации соответствует вершина w орграфа метаинформации, тип набора дуг, исходящих из w , есть «список», и из вершины w выходит дуга D , у которой значение спецификатора принадлежит множеству {«копия»}, «в точности один», то среди дуг, выходящих из вершины v , только одна дуга соответствует D ;

- если вершине v орграфа информации соответствует вершина w орграфа метаинформации, тип набора дуг, исходящих из w , есть «список», и из вершины w выходит дуга D , у которой значение спецификатора принадлежит множеству {«возможное отсутствие», «ноль или один»}, то среди дуг, выходящих из вершины v , может быть одна дуга, которой соответствует D ;

- если вершине v орграфа информации соответствует вершина w орграфа метаинформации, тип набора дуг, исходящих из w , есть «список», и из вершины w выходит дуга D , у которой спецификатор имеет значение «непустой список» или «непустое упорядоченное множество», то среди дуг, выходящих из вершины v , имеется непустое множество дуг $\{d_1, \dots, d_m\}$, каждой из которых соответствует D , конец каждой дуги имеет поле «значение», причем если значение спецификатора есть «непустое множество», то концы попарно различных дуг имеют различные значения поля «значение»;

- если вершине v орграфа информации соответствует вершина w орграфа метаинформации, тип набора дуг, исходящих из w , есть «список», и из вершины w выходит дуга D , у которой спецификатор имеет значение «возможно пустой список» или «возможно пустое упорядоченное множество», то среди дуг, выходящих из вершины v , может быть множество дуг $\{d_1, \dots, d_m\}$, каждой из которых соответствует D , конец каждой дуги имеет поле «значение», причем если значение спецификатора есть «возможно пустое множество», то концы попарно различных дуг имеют различные значения поля «значение»;

- если вершине v орграфа информации соответствует вершина w орграфа метаинформации, тип

набора дуг, исходящих из w , есть «список», и из вершины w выходит дуга D , у которой спецификатор имеет значение «непустая последовательность», то среди дуг, выходящих из вершины v , имеется непустое множество дуг $\{d_1, \dots, d_m\}$, каждой из которых соответствует D , причем конец дуги D имеет поле «сорт» со значением «целое», а конец каждой из дуг d_i , где $i \in [1, m]$, имеет поле «значение», значением которого является константа целого сорта, равная i (концы разных дуг из этого множества имеют попарно различные значения этого поля);

- если вершине v орграфа информации соответствует вершина w орграфа метаинформации, тип набора дуг, исходящих из w , есть «список», и из вершины w выходит дуга D , у которой спецификатор имеет значение «возможно пустая последовательность», то среди дуг, выходящих из вершины v , может быть множество дуг $\{d_1, \dots, d_m\}$, каждой из которых соответствует D , причем конец дуги D имеет поле «сорт» со значением «целое», а конец каждой из дуг d_i , где $i \in [1, m]$, имеет поле «значение», значением которого является константа целого сорта, равная i (концы разных дуг из этого множества имеют попарно различные значения этого поля);

- если вершине v орграфа информации соответствует вершина w орграфа метаинформации и тип набора дуг, исходящих из w , есть «альтернатива», то всем дугам, выходящим из вершины v , соответствует только одна из дуг D , выходящих из вершины w ; если при этом дуга D имеет спецификатор со значением «копия» или «в точности один», то из вершины v выходит только одна дуга d ; если же дуга D имеет спецификатор со значением «непустой список», «непустое упорядоченное множество» или «непустая последовательность», то из вершины v выходит непустое множество дуг, причем если значение спецификатора есть «возможно пустое множество», то концы попарно различных дуг имеют различные значения поля «значение», а если значение спецификатора есть «непустая последовательность», то конец дуги D имеет поле «сорт» со значением «целое», а конец каждой из дуг d_i , где $i \in [1, m]$, имеет поле «значение», значением которого является константа целого сорта, равная i (концы разных дуг из этого множества имеют попарно различные значения этого поля);

- если дуге d орграфа информации соответствует дуга D орграфа метаинформации, у которой ограничитель имеет значение «ссылка», то в вершину, являющуюся концом дуги d , входит не менее одной дуги в орграфе информации.

Пример 1. Соответствие между дугами орграфа на рис. 1 и дугами орграфа на рис. 2, представленными в работе [1], приведено в таблице (дуги метаинформации пронумерованы числами, дуги информации – числами и буквами):

Соответствие между дугами орграфов на рис. 1 и 2 из работы [1].

№	Дуга в метаинформации	Дуги в информации
1	База наблюдений \rightarrow $^{[1]}_{gen}$ Группы	а) База наблюдений в области офтальмологии \rightarrow Группы
2	База наблюдений \rightarrow $^{[1]}_{gen}$ Наблюдения	—
3	Группы \rightarrow $^{[1]}_{gen}$ Группа	а) Группы \rightarrow Главные жалобы; б) Группы \rightarrow Дополнительные жалобы

№	Дуга в метаинформации	Дуги в информации
4	Группа \rightarrow^{+}_{gen} Группы	—
5	Группа $\rightarrow^{[=]}_{all}$ Наблюдения	а) Главные жалобы \rightarrow Наблюдения; б) Дополнительные жалобы \rightarrow Наблюдения
6	Наблюдения \rightarrow^{+}_{gen} Наблюдение	а) Наблюдения \rightarrow Светобоязнь; б) Наблюдения \rightarrow Ощущение инородного тела в глазу; в) Наблюдения \rightarrow Головная боль; д) Наблюдения \rightarrow Температура
7	Наблюдение $\rightarrow^{=}_{all}$ Тип области возможных значений	а) Светобоязнь \rightarrow Тип области возможных значений; б) Ощущение инородного тела в глазу \rightarrow Тип области возможных значений; в) Головная боль \rightarrow Тип области возможных значений; д) Температура \rightarrow Тип области возможных значений
8	Тип области возможных значений $\rightarrow^{=}_{all}$ Составные значения	а,б) Тип области возможных значений \rightarrow Составные значения (2 дуги)
9	Тип области возможных значений $\rightarrow^{=}_{all}$ Качественные значения	а-е) Тип области возможных значений \rightarrow Качественные значения (6 дуг)
10	Тип области возможных значений $\rightarrow^{=}_{all}$ Числовые значения	а) Тип области возможных значений \rightarrow Числовые значения
11	Составные значения \rightarrow^{+}_{all} Характеристика	а,б) Составные значения \rightarrow Присутствие (2 дуги); в,г) Составные значения \rightarrow Характер начала (2 дуги); д) Составные значения \rightarrow Степень проявления; е,ж) Составные значения \rightarrow Периодичность (2 дуги); з) Составные значения \rightarrow Длительность; и) Составные значения \rightarrow Сторона
12	Характеристика $\rightarrow^{=}_{all}$ Тип области возможных значений	а) Присутствие \rightarrow Тип области возможных значений; б) Характер начала \rightarrow Тип области возможных значений; в) Степень проявления \rightarrow Тип области возможных значений; г) Периодичность \rightarrow Тип области возможных значений; д) Длительность \rightarrow Тип области возможных значений; е) Сторона \rightarrow Тип области возможных значений
13	Качественные значения \rightarrow^{+}_{all} Значение	а) Качественные значения \rightarrow <i>отсутствует</i> ; б) Качественные значения \rightarrow <i>имеется</i> ; в) Качественные значения \rightarrow <i>острое</i> ; г) Качественные значения \rightarrow <i>постепенное</i> ; д) Качественные значения \rightarrow <i>слабая</i> ; е) Качественные значения \rightarrow <i>умеренная</i> ; ж) Качественные значения \rightarrow <i>резкая</i> ; з) Качественные значения \rightarrow <i>постоянно</i> ; и) Качественные значения \rightarrow <i>периодически</i> ; к) Качественные значения \rightarrow <i>кратковременная</i> ;

№	Дуга в метаинформации	Дуги в информации
		л) Качественные значения → продолжительная; м) Качественные значения → справа; н) Качественные значения → слева
14	Числовые значения → ^{all} Тип значений	а) Числовые значения → Тип значений
15	Числовые значения → ^[1] _{all} Единица измерения	а) Числовые значения → °С
16	Тип значений → ^{all} Целые значения	—
17	Тип значений → ^{all} Вещественные значения	а) Числовые значения → 34,0 – 42,0

Будем говорить, что оргграф информации является формально полным относительно оргграфа метаинформации, если оргграф информации соответствует этому оргграфу метаинформации.

ПОРОЖДЕНИЕ ОРГРАФА ИНФОРМАЦИИ ПО ОРГРАФУ МЕТАИНФОРМАЦИИ

В оргграфовой связанной двухуровневой модели анкеты редактор информации порождает оргграф информации по оргграфу метаинформации с помощью следующего порождающего процесса, в котором пользователь разрешает неопределенности, возникающие на каждом шаге процесса порождения (выбирает ответы из предлагаемых списков, либо их формулирует):

- на первом шаге процесса порождается начальная вершина v оргграфа информации, значением поля «значение» которой является порождаемая константа c^* сорта «строковое»;
- если вершине v оргграфа информации соответствует вершина w оргграфа метаинформации и вершина w имеет поля «имя» и «сорт» со значением s , то для вершины v порождается поле «значение» со значением c (константой сорта s);
- если вершине v оргграфа информации соответствует вершина w оргграфа метаинформации и вершина w имеет поле «значение» со значением c , то для вершины v порождается поле «значение» со значением c ;
- если вершине v оргграфа информации соответствует вершина w оргграфа метаинформации, тип набора дуг, исходящих из w , есть «список», и из вершины w выходит дуга D , у которой спецификатор имеет значение, принадлежащее множеству {«копия», «в точности один»}, то в оргграфе информации порождается дуга, выходящая из вершины v ;
- если вершине v оргграфа информации соответствует вершина w оргграфа метаинформации, тип набора дуг, исходящих из w , есть «список», и из вершины w выходит дуга D , у которой спецификатор имеет значение, принадлежащее множеству {«возможное отсутствие», «ноль или один»}, то в оргграфе информации может порождаться дуга, выходящая из вершины v ;
- если вершине v оргграфа информации соответствует вершина w оргграфа метаинформации, тип набора дуг, исходящих из w , есть «список», и из вершины w выходит дуга D , у которой спецификатор имеет значение «непустой список» или «непустое

упорядоченное множество», то в оргграфе информации порождается непустое множество дуг $\{d_1, \dots, d_m\}$, выходящих из вершины v , каждой из которых соответствует D , конец каждой дуги имеет поле «значение», причем если признак дуги D имеет значение «непустое упорядоченное множество», то значения этих полей «значение» попарно различны;

- если вершине v оргграфа информации соответствует вершина w оргграфа метаинформации, тип набора дуг, исходящих из w , есть «список», и из вершины w выходит дуга D , у которой спецификатор имеет значение «возможно пустой список» или «возможно пустое упорядоченное множество», то в оргграфе информации может порождаться непустое множество дуг $\{d_1, \dots, d_m\}$, выходящих из вершины v , каждой из которых соответствует D , конец каждой дуги имеет поле «значение», причем если признак дуги D имеет значение «возможно пустое упорядоченное множество», то значения этих полей «значение» попарно различны;
- если вершине v оргграфа информации соответствует вершина w оргграфа метаинформации, тип набора дуг, исходящих из w , есть «список», и из вершины w выходит дуга D , у которой спецификатор имеет значение «непустая последовательность», а конец имеет поле «сорт» со значением «целое», то в оргграфе информации порождается непустое множество дуг $\{d_1, \dots, d_m\}$, выходящих из вершины v , каждая из которых соответствует D , а конец каждой из дуг d_i , где $i \in [1, m]$ имеет поле «значение», в качестве значения которого порождается константа целого сорта, равная i (концы разных дуг из этого множества имеют попарно различные значения этого атрибута);
- если вершине v оргграфа информации соответствует вершина w оргграфа метаинформации, тип набора дуг, исходящих из w , есть «список», и из вершины w выходит дуга D , у которой спецификатор имеет значение «возможно пустая последовательность», а конец имеет поле «сорт» со значением «целое», то в оргграфе информации может порождаться непустое множество дуг $\{d_1, \dots, d_m\}$, выходящих из вершины v , каждая из которых соответствует D , а конец каждой из дуг d_i , где $i \in [1, m]$ имеет поле «значение», в качестве значения которого порождается константа целого сорта равная i (концы разных дуг из этого множества имеют попарно различные значения этого атрибута);
- если вершине v оргграфа информации соответствует вершина w оргграфа метаинформации, тип

набора дуг, исходящих из w , есть «альтернатива», то порождается одна или несколько дуг, выходящих из вершины v , которые соответствуют только одной из дуг D , выходящих из вершины w ; если при этом дуга D имеет спецификатор со значением «копия» или «в точности один», то порождается только одна дуга d , выходящая из вершины v ; если же дуга D имеет спецификатор со значением «непустой список», «непустое упорядоченное множество» или «непуста последовательность», то порождается непустое множество дуг, выходящих из вершины v ;

- если дуга d в орграфе информации порождается по дуге D орграфа метайнформации, у которой ограничитель имеет значение «ссылка», то концом дуги d является уже существующая в орграфе информации вершина (при этом, за исключением случая, когда эта вершина является начальной, она всегда является концом еще одной дуги в орграфе информации); если ограничитель имеет значение «порождение», то концом дуги d является новая вершина; если же ограничитель имеет значение «порождение или ссылка», то возможен любой из этих случаев.

Пример 2. Процесс порождения орграфа информации по орграфу метайнформации, приведенному на рис. 1 в работе [1], может быть представлен следующим образом:

По дуге 1 порождается дуга 1а: $\{1 - 1а\}$, по дуге 3 порождается дуга 3а: $\{3 - 3а\}$ и так далее – согласно порядку дуг в орграфе метайнформации и порядку дуг в орграфе информации, соответствующих дуге из орграфа метайнформации. Финальным является порождение $\{17 - 17а\}$.

Заметим, что в процессе порождения орграфа информации по орграфу метайнформации, орграф информации может находиться в состояниях, в которых он не является формально полным относительно этого орграфа метайнформации.

Пример 3. Состояние орграфа информации не является формально полным на всех шагах порождающего процесса примера 2, кроме двух, и является формально полным после выполнения шагов порождения $\{13 - 13м\}$, $\{13 - 13н\}$ и $\{17 - 17а\}$.

ДВА ОСОБЫХ СЛУЧАЯ В СООТВЕТСТВИИ ОРГРАФОВ ИНФОРМАЦИИ И МЕТАИНФОРМАЦИИ

До сих пор рассматривался случай, когда дана некоторая система понятий, представленная орграфом метайнформации в орграфовой связанной двухуровневой модели информационных единиц, и по ней порождается система экземпляров понятий, представленная орграфом информации в этой модели.

Однако в некоторых случаях из заданной вершины орграфа информации необходимо:

- породить произвольный ориентированный подграф g' , для которого в метайнформации нет подграфа G , которому g' соответствует;

- сделать ссылку на любую вершину произвольного орграфа информации или метайнформации g'' .

Для обеспечения таких возможностей в орграф метайнформации следует ввести специальный тип вершин (будем обозначать их символом Λ), а соот-

ветствие между орграфами информации и метайнформации дополнить следующими условиями:

- если вершина v_0 орграфа информации ig соответствует вершине w_0 орграфа метайнформации MG , являющейся начальной вершиной для дуг D_1, \dots, D_n ($n \geq 0$) и $D'(w_0 \rightarrow^{D'} \Lambda)^1$, то вершина v_0 является, с одной стороны, начальной вершиной подграфа g орграфа информации ig , определяемого правилами соответствия между орграфами информации и метайнформации, описанными в разделе 6, а с другой стороны – начальной вершиной подграфа g' орграфа информации ig – произвольного орграфа, представленного на языке метайнформации (соответственно этот подграф ничему не соответствует в орграфе метайнформации MG);

- если дуга d ($v_0 \rightarrow^d v$) орграфа информации ig соответствует дуге D' ($w_0 \rightarrow^{D'} \Lambda$) орграфа метайнформации MG^2 , то подграф g'' орграфа информации ig с начальной вершиной v является орграфом информации с произвольной метайнформацией, либо произвольным орграфом, представленным на языке метайнформации.

Пример 4. На рис. 1 представлен орграф метайнформации, описывающий структуру шаблонов сообщений, посредством которых агенты облачной платформы IASaaS взаимодействуют друг с другом [13–15].

С одной стороны, такая структура фиксирует то, что каждый шаблон сообщений имеет внутреннее имя, содержательное описание назначения (на естественном языке) и, возможно, исполняемый код шаблона сообщений (байт-код), представляющий его процедурную часть. С другой стороны, поскольку шаблон сообщений определяет, по существу, язык представления некоторого множества сообщений, такая структура должна позволять для каждого шаблона сообщений описать структуру (синтаксис языка представления) содержательной информации, передаваемой в сообщениях, формируемых по этому шаблону (содержимого сообщений) – в случае, когда шаблон сообщений описывает не просто сообщения-команды, а сообщения, которые должны содержать некоторые данные. Структура содержимого сообщений в каждом шаблоне сообщений описывается произвольным орграфом, представленным на языке метайнформации. Начальной вершиной этого орграфа является вершина «Шаблон Запрос от агента Вид». Спецификатор дуги «пропуск» имеет символическое обозначение «~».

Пример 5. На рис. 2 представлен орграф метайнформации, описывающий структуру технологии разработки интегрированных интеллектуальных решателей задач (ИИРЗ). Эта технология (как понятие) имеет два атрибута, значениями которых являются компоненты ИИРЗ (онтология базы знаний и решатель).

¹ Значение спецификатора дуги D' в этом случае есть «пропуск».

² Значение ограничителя дуги D' в этом случае есть «ссылка», а значением спецификатора могут быть только «в точности один», «непустой список», «непустое упорядоченное множество», «ноль или один», «возможно пустой список» и «возможно пустое упорядоченное множество».

тель задач). Каждый атрибут имеет также ссылку на работу, с помощью которой создается соответствующая компонента. Решатель задач имеет фиксированную структуру, которая представлена соответствующим орграфом метаинформации, хранящимся в

фонде платформы IASaaS; на эту метаинформацию и ссылается соответствующий атрибут. Онтология базы знаний является произвольной метаинформацией, поэтому значением атрибута «Онтология базы знаний» является символ «Λ».

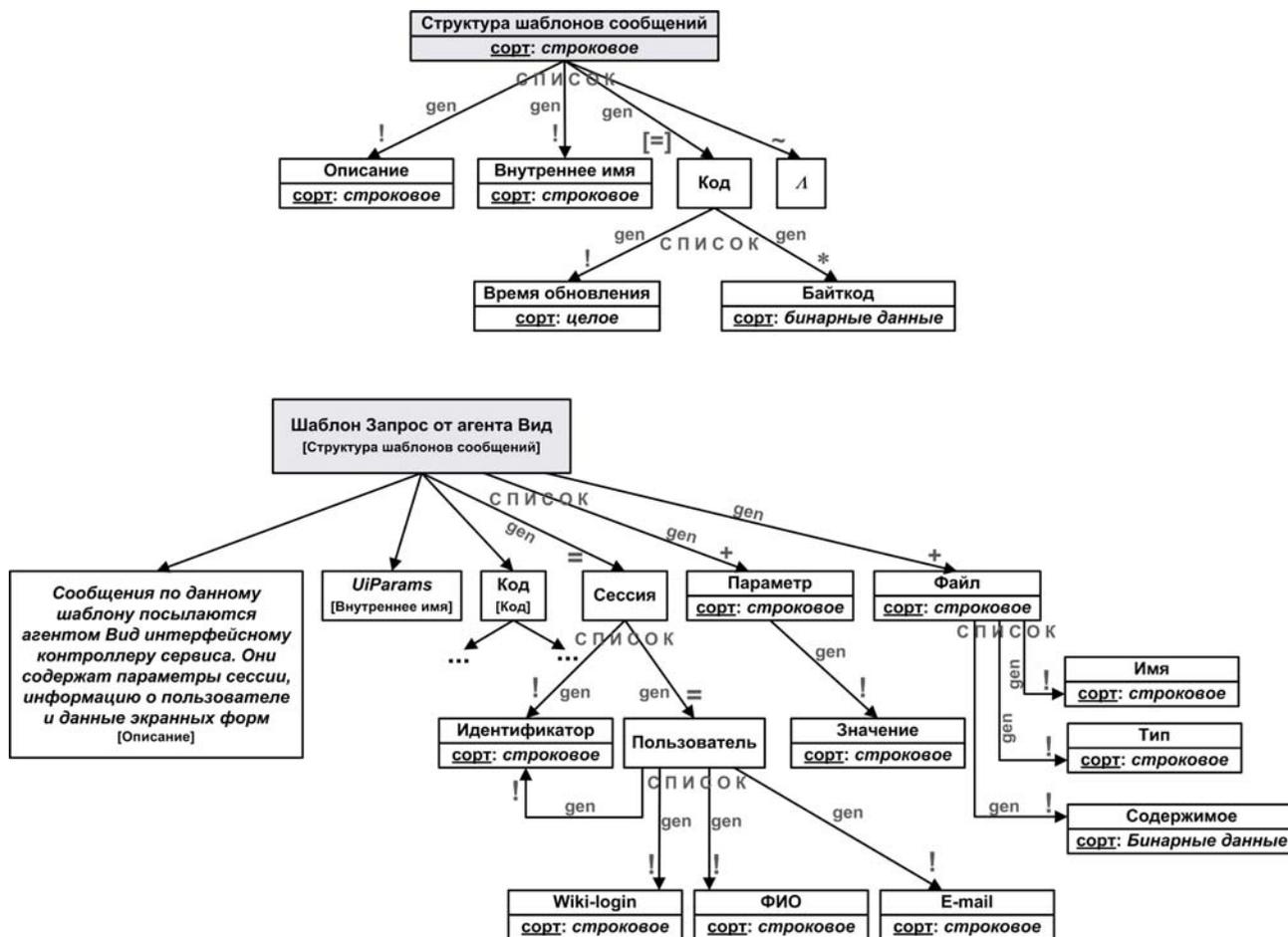


Рис. 1. Орграф метаинформации, описывающий структуру шаблонов сообщений платформы IASaaS, и орграф информации, описывающий устройство шаблона сообщений «Шаблон Запрос от агента Вид»

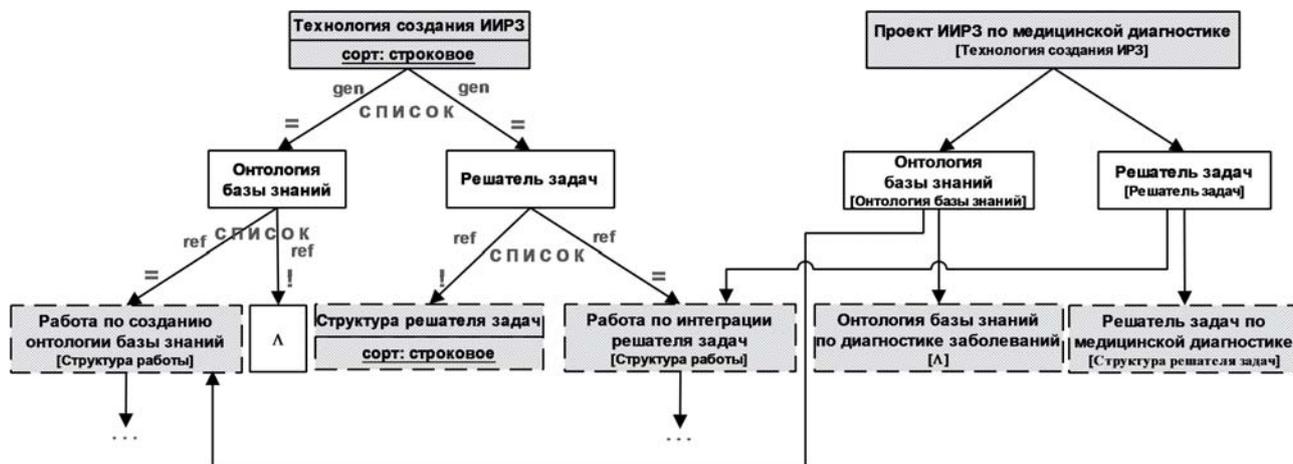


Рис. 2. Орграф метаинформации, описывающий структуру технологии по разработке интегрированных интеллектуальных решателей задач, и орграф порожденной по нему информации, описывающий проект такого решателя

По метаинформации, представляющей технологию, порождается информация, представляющая проект конкретного ИИРЗ (в данном случае ИИРЗ медицинской диагностики). Экземпляр понятия «Проект ИИРЗ по медицинской диагностике» имеет те же два атрибута. Значением атрибута «Онтология базы знаний» является произвольный орграф метаинформации, создаваемый при работе над проектом ИИРЗ, а значением атрибута «Решатель задач» – ссылка на решатель задач по медицинской диагностике, порождаемый по метаинформации «Структура решателя задач».

Зеркальным является случай, когда из вершины орграфа метаинформации выходит дуга, которая входит в вершину орграфа информации. В этом случае действует следующее условие на соответствие между орграфами информации и метаинформации: если из вершины w орграфа метаинформации выходит дуга D' , которая входит в вершину v орграфа информации³, и в орграфе информации вершина v_0 соответствует вершине w , то в орграфе информации из вершины v_0 выходит дуга d' , которая входит в вершину v и соответствует дуге D' орграфа метаинформации.

Пример 6. Продолжая описание рис. 2, отметим, что разработка онтологии базы знаний должна выполняться согласно работе по созданию онтологии базы знаний, описанной в соответствующем орграфе информации, а разработка интегрированного решателя задач – согласно работе по созданию интегрированных решателей задач (которая также описывается в соответствующем орграфе информации). В метаинформации это задаётся дугами «Онтология базы знаний \rightarrow Работа по созданию онтологии базы знаний» и «Решатель задач \rightarrow Работа по интеграции решателя задач» соответственно. При разработке интегрированного решателя задач по такой технологии в порождаемом информационном ресурсе по этим дугам порождаются, соответственно, следующие: «Онтология базы знаний \rightarrow Работа по созданию онтологии базы знаний» и «Решатель задач \rightarrow Работа по интеграции решателя задач», концами которых являются те же вершины, что находятся на концах дуг из метаинформации.

Для этих двух особых случаев (см. Примеры 5 и 6) могут быть сформулированы и правила порождения орграфа информации по орграфу метаинформации.

КЛОНИРОВАНИЕ ВЕРШИН, МНОЖЕСТВ ВЕРШИН И ОРГРАФОВ ИНФОРМАЦИИ

Порождение вершин и дуг орграфа информации по вершинам и дугам орграфа метаинформации – это не единственный способ построения орграфа информации. Другим способом является клонирование вершин и дуг некоторого орграфа информации g_1 при построении другого орграфа информации g_2 (при этом соответствие между орграфом g_2 и его орграфом метаинформации G_2 должно выполняться). Обозначим v' вершину орграфа информации g_1 , а v – вершину орграфа информации g_2 , которая должна

быть получена клонированием v' . Клонирование может быть независимым и зависимым.

При независимом клонировании вершины в орграфе информации g_2 создается вершина v , у которой:

- значения всех полей совпадают со значениями одноименных полей вершины v' ;
- дуги, выходящие из v , являются клонами дуг, выходящих из v' .

В дальнейшем изменения значений полей и набора выходящих дуг для этих вершин, никак не связаны между собой.

При независимом клонировании множества вершин в орграфе информации g_2 создается множество вершин v_1, \dots, v_n , количество которых совпадает с количеством вершин в клонируемом множестве, такое, что каждый его элемент является независимым клоном соответствующей вершины из орграфа информации g_1 .

Клонирование орграфа информации g , не имеющего вершин, одновременно принадлежащих и какому-либо другому орграфу информации g' , можно рассматривать как независимое клонирование множества вершин орграфа g . Клонирование орграфа информации g , имеющего вершины v_1, \dots, v_n , одновременно принадлежащие некоторым другим орграфам информации g_1, \dots, g_m , в которые входят дуги d_1, \dots, d_k , выходящие из вершин орграфа g , имеет два варианта:

- поверхностное клонирование, при котором орграфы g_1, \dots, g_m не клонируются, а клоны дуг d_1, \dots, d_k входят в соответствующие вершины v_1, \dots, v_n ;
- глубокое клонирование, при котором клонируются и все орграфы g_1, \dots, g_m .

Пример 7. Рассмотрим процесс создания резервной копии базы знаний о заболеваниях. Такая база содержит ссылки на базу наблюдений – на признаки и их значения. На рис. 3 представлены слева направо: метаинформация базы наблюдений [16], сформированная на её основе база наблюдений в области офтальмологии [17], исходная база знаний – база знаний о заболеваниях в области офтальмологии и её основа – метаинформация базы знаний о заболеваниях [16, 18] (со ссылками на понятия метаинформации базы наблюдений, по которым в порождаемой информации также необходимо порождать ссылки на информацию, описывающую наблюдения). Спецификатор дуги «непустая последовательность» на нем имеет символьное обозначение «^». Со временем знания эксперта расширяются, и он стремится модифицировать свою базу знаний о заболеваниях, при этом логичным является сохранение резервной отлаженной копии исходной базы знаний. При создании копии такой базы знаний создаётся информация, приведённая в правой части рис. 3. Она строится по той же метаинформации (База знаний о заболеваниях); её начальная вершина есть имя, отличное от имени исходной базы («Копия базы знаний в области офтальмологии»); вершины, которые принадлежали исходной базе знаний, имеют свои независимые клоны в копии с такими же полями и исходящими дугами («Клинические проявления», «Клиническое проявление для признака Головная боль», 1, 1, 1 - 10; к вершинам из орграфа базы наблюдений в области офтальмологии «Головная боль» и «имеется» (в Базе наблюдений) идут дуги, так же как в исходной Базе знаний в области офтальмологии.

³ Значение ограничителя дуги D' в этом случае есть «ссылка», а значением спецификатора могут быть только «копия» и «возможное отсутствие».

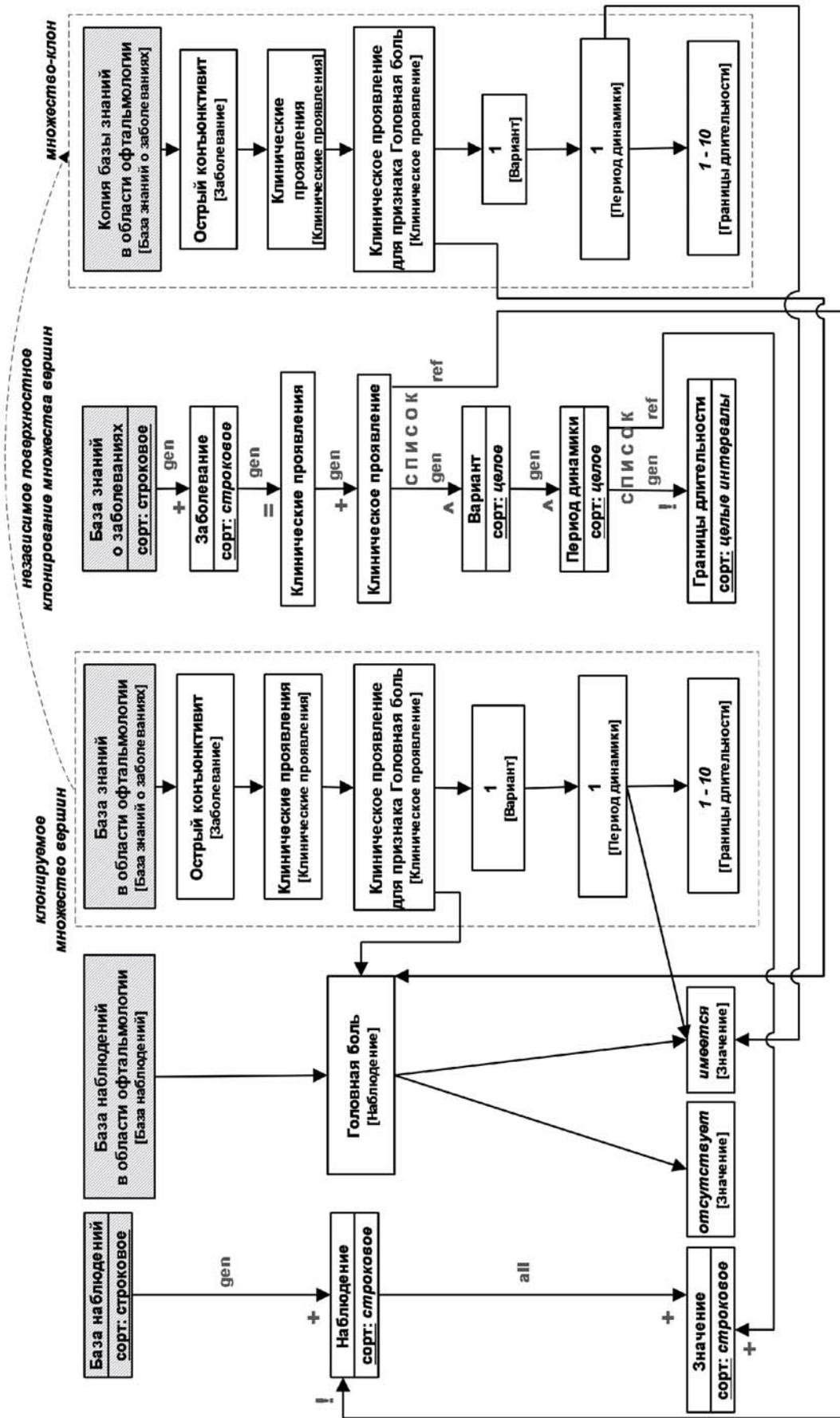


Рис. 3. Независимое поверхностное клонирование базы знаний о заболеваниях в области офтальмологии

Пример 8. Рассмотрим процесс передачи информации от одного эксперта другому на примере медицинской диагностики в области офтальмологии. Первый эксперт создаёт базу наблюдений и базу знаний и наполняет их, приводя к некоторому качественно-полезному состоянию (тривиальный пример см. на рис. 4). Далее ему требуется передать эту пару баз знаний другому эксперту, для того чтобы тот начал их использовать в своих целях. При выполнении глубокого клонирования этой информации создается информация с начальной вершиной «Копия базы наблюдений в области офтальмологии» и информация с начальной вершиной «Копия базы знаний в области офтальмологии». Их метainформации – те же что и у исходной информации (соответственно). В информации «Копия наблюдений в области офтальмологии» создаются вершины-клоны для вершин «Головная боль», «отсутствует», «имеется». В информации «Копия базы знаний в области офтальмологии» создаются клоны исходных вершин «Острый конъюнктивит», «Клинические проявления», «Клиническое проявление для признака Головная боль», 1, 1, 1-10. А также от вершины «Клиническое проявление для признака Головная боль» создаётся дуга к вершине-клону «Головная боль» (находящейся в орграфе «Копия базы наблюдений в области офтальмологии») и от вершины 1 (соответствующей вершине в орграфе метainформации «Период динамики») создаётся дуга к вершине-клону «имеется».

При зависимом клонировании вершины v' в орграфе информации g_2 создается вершина v , у которой

значение поля «значение» совпадает со значением этого поля у клонируемой вершины v' . В дальнейшем при редактировании:

- изменение значения поля «значение» вершины v' приводит к соответствующему изменению значения этого поля у всех вершин-клонов;
- изменение значения поля «значение» вершины v невозможно;
- удаление вершины v' приводит к удалению всех ее клонов;
- удаление любого клона вершины v' допустимо и никак не отражается на вершине v' и других ее клонах.

При зависимом клонировании множества вершин в орграфе информации g_2 создается множество вершин v_1, \dots, v_n , количество которых совпадает с количеством вершин в клонируемом множестве такое, что каждый его элемент является зависимым клоном соответствующей вершины из орграфа информации g_1 . В дальнейшем при редактировании:

- при удалении некоторой вершины из клонируемого множества вершин все ее клоны должны быть удалены из каждого множества-клона;
- при добавлении некоторой вершины в клонируемое множество вершин во все клоны этого множества должны быть добавлены клоны добавленной вершины;
- ни в одно из множеств-клонов не могут быть добавлены новые вершины и ни из одного из множеств-клонов не могут быть удалены вершины.

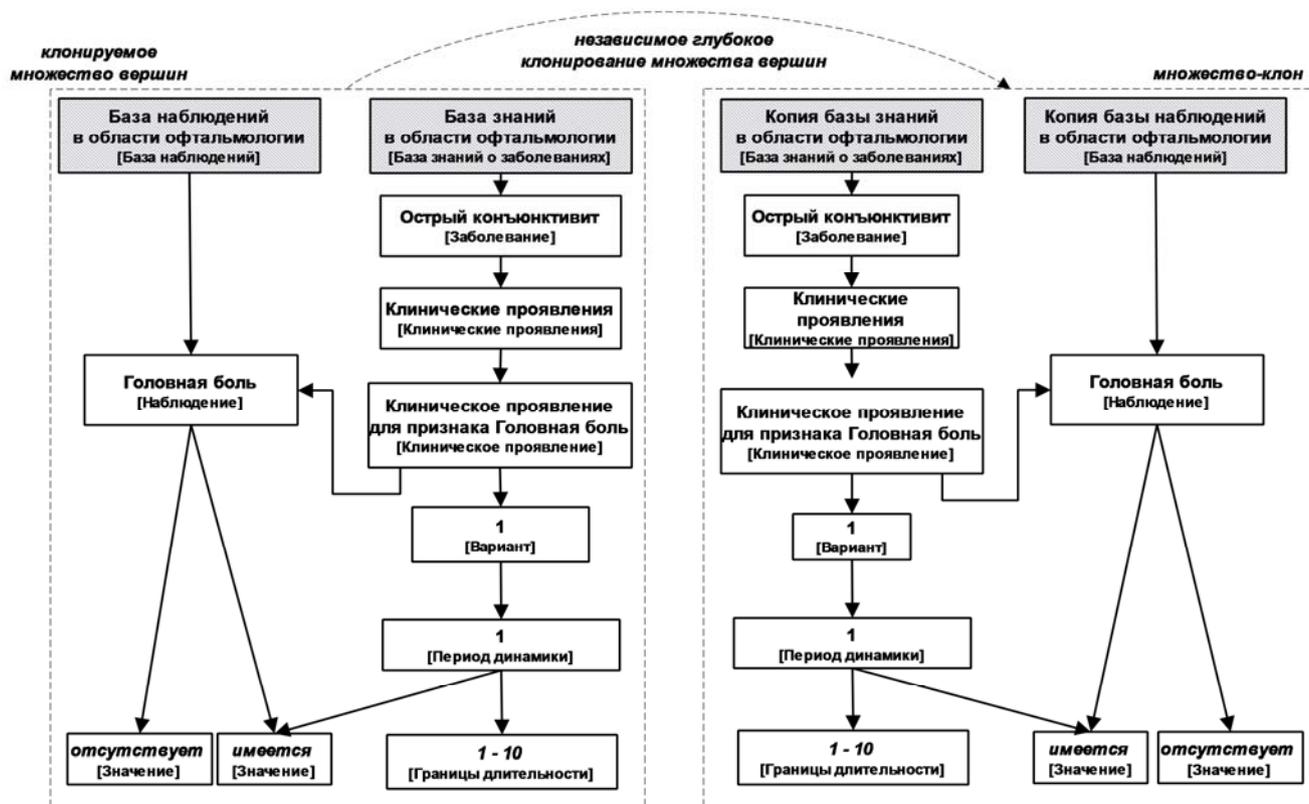


Рис. 4. Независимое глубокое клонирование базы наблюдений и базы знаний о заболеваниях в области офтальмологии

Пример 9. В процессе контекстно-зависимого порождения орграфа информации g_2 , описывающего простую базу знаний о нормальных значениях в области офтальмологии, по орграфу метаинформации G_2 , описывающему простые базы знаний о нормальных значениях в различных разделах медицины (рис. 5), с использованием орграфа информации g_1 (орграф его метаинформации G_1 , описывающий простые базы наблюдений в различных разделах медицины, представлен на рис. 6), описывающего простую базу наблюдений в области офтальмологии, в качестве контекста требуется, чтобы (рис. 7):

1) в g_2 множество потомков вершины «Простая база знаний о нормальных значениях: офтальмология», которой соответствует вершина «Простая база знаний о нормальных значениях» в орграфе G_2 , было

результатом зависимого клонирования множества вершин – потомков вершины «Простая база наблюдений: офтальмология» в орграфе g_1 , которой соответствует вершина «Простая база наблюдений» в орграфе метаинформации G_1 ;

2) в g_2 для каждой вершины со значением p^* , которой в G_2 соответствует вершина «признак», множество потомков было результатом зависимого клонирования элементов непустого подмножества вершин – потомков вершины в g_1 со значением p^* , которой в G_1 соответствует вершина «наблюдение».

Отношение между вершинами и множествами вершин орграфа информации и их зависимыми клонатами аналогично отношению между орграфами метаинформации и порожденными по ним орграфами информации.

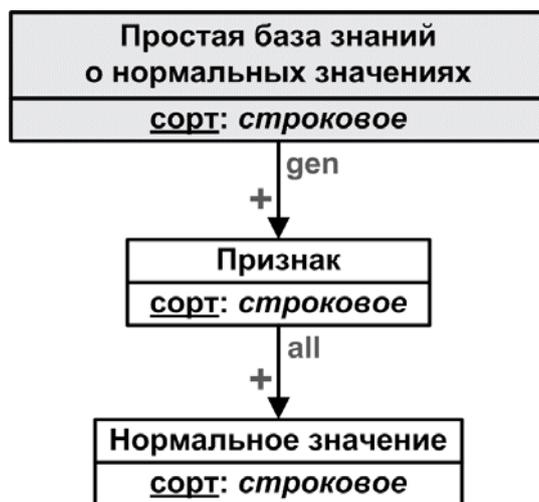


Рис. 5. Орграф метаинформации для простых баз знаний о нормальных значениях в различных разделах медицины

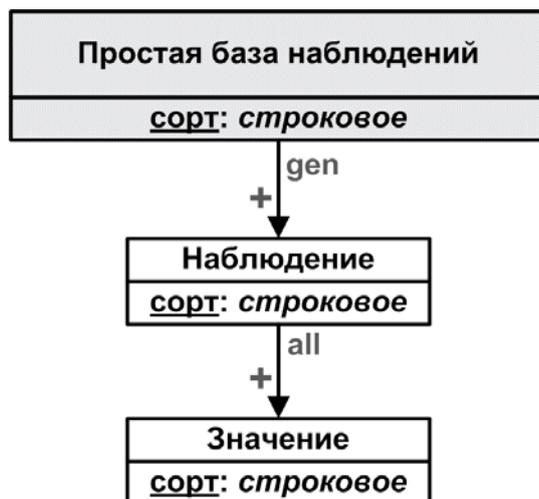


Рис. 6. Орграф метаинформации для простых баз наблюдений в различных разделах медицины

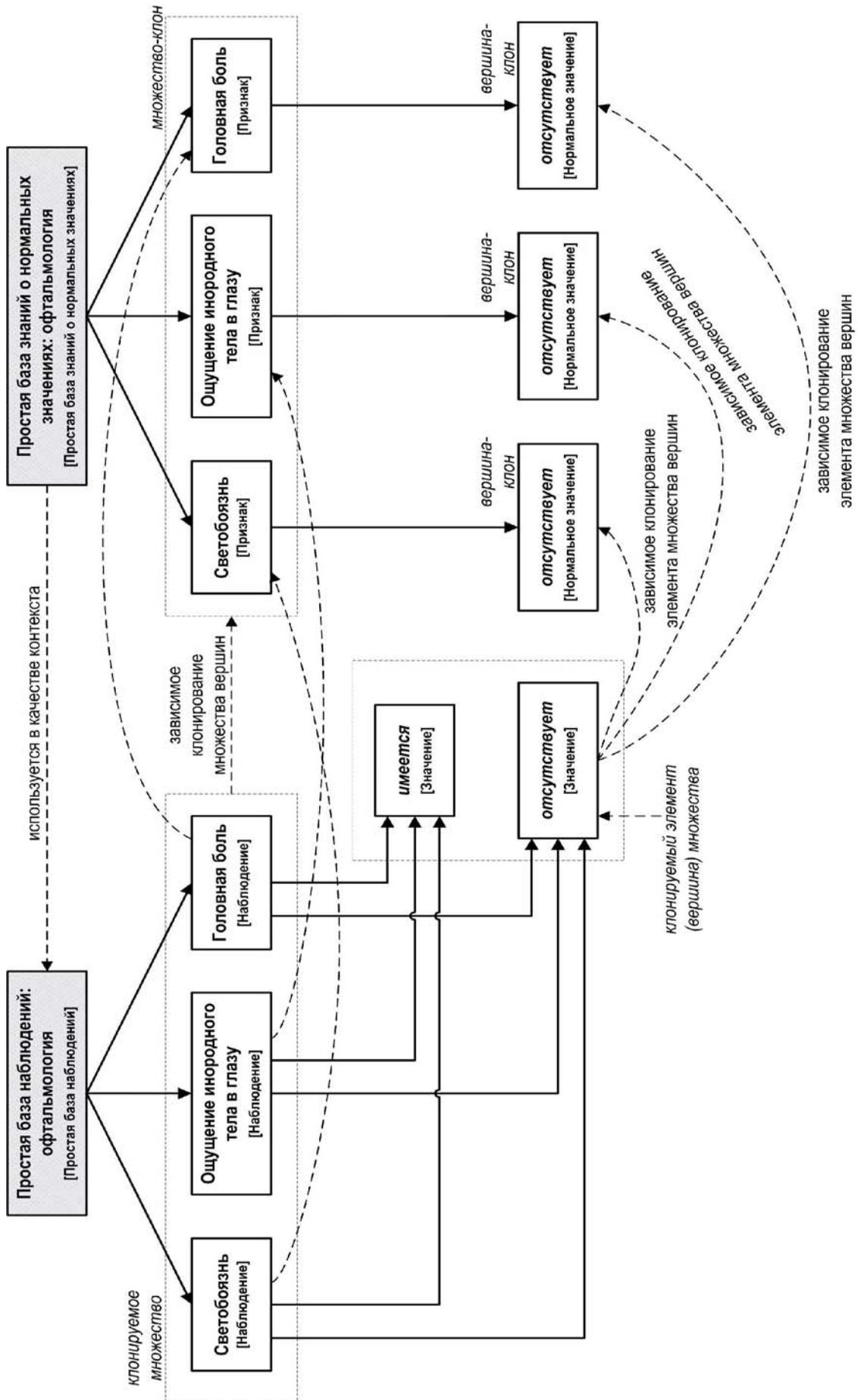


Рис. 7. Контекстно-зависимое порождение орграфа информации, описывающего простую базу знаний о нормальных значениях в области офтальмологии

ЗАКЛЮЧЕНИЕ

В настоящей статье рассмотрена модель процесса порождения орграфов, представляющих сложно-структурированную информацию, на основе орграфовой связной двухуровневой модели информационных единиц и модели соответствия между орграфами информации и метаинформации. Заложенные в модель особые случаи соответствия орграфов информации и метаинформации позволяют формировать иерархическое многоуровневое (число уровней больше двух) представление информации.

В качестве еще одного (помимо порождения по метаинформации) способа построения орграфа информации рассматривается зависимое и независимое клонирование вершин и дуг некоторого другого орграфа информации (при этом сохраняется соответствие между формируемым таким образом орграфом информации и орграфом его метаинформации). Механизм зависимого клонирования направлен на поддержку контекстно-зависимого формирования орграфов информации.

СПИСОК ЛИТЕРАТУРЫ

1. Грибова В.В., Клещев А.С., Москаленко Ф.М., Тимченко В.А. Двухуровневая модель сложноструктурированных информационных единиц, соответствующая метафоре анкетирования // Научно-техническая информация. Сер. 2. – 2015. – № 10. – С. 1-10.
2. Neologism: Easy Vocabulary Publishing. – URL: <http://ceur-ws.org/Vol-368/paper10.pdf> (дата обращения: 10.07.2015).
3. Dragoni M., Ghidini C., Bosca A. Multilingual MoKi: How to Manage Multilingual Ontologies in a Wiki / ed. Philipp Cimiano, Miriam Fernández, Vanessa Lopez, Stefan Schlobach, Johanna Völker // ESWC (Satellite Events). – Springer. – 2013. – P. 162-166.
4. Guoxing Fu, Weinian Rao. An introduction of Knoodl. – URL: <http://tinman.cs.gsu.edu/~raj/8711/sp11/presentations/knoodlReport.pdf> (дата обращения 10.07.2015).
5. OntoWiki: a tool providing support for agile, distributed knowledge engineering scenarios. – URL: <http://aksw.org/Projects/OntoWiki.html> (дата обращения 10.07.2015).
6. SOBOLEO – Editor and Repository for Living Ontologies. – URL: <http://knowledge-maturing.com/services-and-tools/tools/> (дата обращения 10.07.2015).
7. PoolParty – thesaurus management system and a SKOS editor for the Semantic Web. – URL: <http://www.poolparty.biz/> (дата обращения 10.07.2015).
8. The Protégé Ontology Editor and Knowledge Acquisition System. – URL: <http://protege.stanford.edu/> (дата обращения 07.10.2015).
9. Kleppe A., Warmer S., Bast W. MDA Explained. The Model Driven Architecture: Practice and Promise. – Addison-Wesley, 2003. – 192 p.
10. Varro D., Pataricza A. VPM: A visual, precise and multilevel metamodeling framework for describing mathematical domains and UML // Journal of Software and Systems Modeling. – 2003. – Vol. 2(3). – P. 187-210.
11. Muller P.-A., Fleurey F., Jezequel J.-M. Weaving Executability into Object-Oriented Meta-languages / ed. L. S.K. Briand // MoDELS 2005. LNCS. – Springer. – 2005. – Vol. 3713. – P. 264 – 278.
12. MDA web site. – URL: <http://www.omg.org/mda> (дата обращения: 18.02.2015).
13. Грибова В.В., Клещев А.С., Крылов Д.А., Москаленко Ф.М., Смагин С.В., Тимченко В.А., Тютюнник М.Б., Шалфеева Е.А. Проект IACPaas. Комплекс для интеллектуальных систем на основе облачных вычислений // Искусственный интеллект и принятие решений. – 2011. – №1. – С. 27-35.
14. Gribova V.V., Kleschev A.S., Krylov D.A., Moskalenko Ph.M., Timchenko V.A., Shalfeeva E.A., Goldstein M.L. A software platform for the development of intelligent multi-agent internet-services // Proceedings of the Distributed Intelligent Systems and Technologies Workshop (DIST'2013). 1-4 July 2013. St. Petersburg, Russia. – СПб: Politechnika-service. – 2015. – P. 29-36.
15. Крылов Д.А., Москаленко Ф.М., Тимченко В.А. Технология разработки решателей задач интеллектуальных систем с использованием инструментальных сервисов облачной платформы IACPaas // Материалы V Международной научно-технической конференции OSTIS-2015. Минск, 19–21 февраля 2015 г. – Минск: БГУИР. – 2015. – С. 193-198.
16. Клещев А.С., Москаленко Ф.М., Черняховская М.Ю. Модель онтологии предметной области «Медицинская диагностика». Часть 1. Неформальное описание и определение базовых терминов // Научно-техническая информация. Сер. 2. – 2005. – №12. – С.1-7.
17. Черняховская М.Ю., Москаленко Ф.М. База наблюдений в области офтальмологии для банка медицинских знаний // Информатика и системы управления. – 2009. – № 2(20). – С. 40-49.
18. Клещев А.С., Москаленко Ф.М., Черняховская М.Ю. Модель онтологии предметной области «Медицинская диагностика». Часть 2. Формальное описание причинно-следственных связей, причин значений признаков и причин заболеваний // Научно-техническая информация. Сер. 2. – 2006. – №2. – С.19-30.

Материал поступил в редакцию 23.07.15.

Сведения об авторах

ГРИБОВА Валерия Викторовна – доктор технических наук, заместитель директора по научной работе, старший научный сотрудник Института автоматизации и процессов управления Дальневосточного отделения Российской академии наук (ИАПУ ДВО РАН), г. Владивосток
e-mail: gribova@iacp.dvo.ru

КЛЕЩЕВ Александр Сергеевич – доктор физико-математических наук, профессор, главный научный сотрудник ИАПУ ДВО РАН, г. Владивосток
e-mail: kleshev@iacp.dvo.ru

МОСКАЛЕНКО Филипп Михайлович – кандидат технических наук, научный сотрудник ИАПУ ДВО РАН, г. Владивосток
e-mail: philipmm@iacp.dvo.ru

ТИМЧЕНКО Вадим Андреевич – кандидат технических наук, научный сотрудник ИАПУ ДВО РАН, г. Владивосток
e-mail: vadim@dvo.ru

Указатель статей, опубликованных в сборнике «Научно-техническая информация», и Авторский указатель за 2015 год*

Указатель статей

МАТЕРИАЛЫ ВСЕРОССИЙСКОЙ НАУЧНО- ПРАКТИЧЕСКОЙ КОНФЕРЕНЦИИ «СОВРЕМЕННЫЕ ПРОБЛЕМЫ ФУНДАМЕНТАЛЬНЫХ НАУК В БИБЛИОГРАФИЧЕСКИХ КЛАССИФИКАЦИОННЫХ СИСТЕМАХ» Москва, ВИНТИ РАН, октябрь 2014 г.

Антопольский А.Б., Белоозеров В.Н., Маркарова Т.С., Дмитриева Е.Ю. Установление соответствий рубрик ГРНТИ рубрикам других систем класси- фикации научной и технической инфор- мации	3 (1) 3*	Белоозеров В.Н., Шабурова Н.Н. Сопоставление библиографических клас- сификаций по полупроводникам и нано- технологиям в тезаурусном формате	3 (1) 47
Саркисян Д.Б., Антошкова О.А., Астахова Т.С. Универсальная десятич- ная классификация – международная классификационная система	3 (1) 19	Ивановский А.А. Опыт БЕН РАН по со- поставлению различных классификаци- онных систем	3 (1) 52
Соловьёва И. М., Дмитриева Е.Ю. Опыт работы с классификационными системами в ВИНТИ РАН	3 (1) 23	Шабурова Н.Н. Использование класси- фикационных систем в библиотеках Сибирского отделения Российской ака- демии наук	3 (1) 55
Ефременкова В.М., Смирнова О.В. Использование УДК и ГРНТИ для оценки объемов информационно-библиотечных массивов документов	3 (1) 27	Гоннова С. М., Никольская И. Ю., Антошкова О. А. О Концепции разработ- ки системы соответствий между классифи- кациями научно-технической информации	3 (1) 57
Смыслова И.С., Соловьёва И.М. Классификационные системы – Государственный рубрикатор научно- технической информации и Рубрикатор ВИНТИ	3 (1) 32	Калёнов Н. Е., Белоозеров В.Н. Форми- рование терминологических словарей по лексике классификационных систем	3 (1) 60
Смирнова О.В. Методика и практика со- ставления индексов УДК	3 (1) 35	Верная Л. А. Физика конденсированных сред в УДК, ГРНТИ и PACS (к вопросу о построении таблиц соответствий индексов в различных системах классификаций)	3 (1) 70
Антошкова О.А., Астахова Т.С., Дмитриева Е.Ю., Смирнова О.В., Сурикова Н.Г. Методика сопоставления УДК с «Рубрикатором по нанотехно- логиям и наноматериалам»	3 (1) 37	ОБЩИЙ РАЗДЕЛ	
Смирнова О.В. Автоматизированная система ведения УДК в ВИНТИ РАН	3 (1) 40	Семенюк Э.П. Глобализация информа- ционного пространства и человечество	1 (1) 1
Антошкова О.А. О новом учебном посо- бии по Универсальной десятичной клас- сификации	3 (1) 43	Сютюренко О.В. Цифровая среда: тренды и риски развития	2 (1) 1
		Астахова Л.В. Информационная безо- пасность: риски, связанные с культурным капиталом персонала	4 (1) 1
		Плешкевич Е.А. Дискуссия о природе информации и путях построения ее фило- софской концепции (Обзор)	4 (1) 14
		Калёнов Н.Е. Об информационном со- провождении фундаментальных научных исследований	4 (2) 1
		Ермолаева А.В. Психологический фактор в организации информационно-документа- ционного обеспечения управления	5 (1) 1
		Брежнева В.В., Гиляревский Р.С. От информационного обслуживания к ин- формационному менеджменту	5 (1) 7

* 3 – означает номер сборника, (1) – серию, 3 – страницу

Родионов И.И., Цветкова В.А. К вопросу об информационном менеджменте и информатике	5 (1) 10	Карпюк Д.В. Информационные ресурсы социальных сетей: возможности депонирования	4 (1) 19
Двоеносова Г.А. Феномен документа	6 (1) 1	Раевская Е.Г. Деятельность Китайской академии наук и отражение ее научно-технических публикаций в международных информационных системах	4 (1) 24
Зибарева И.В. Информационные ресурсы по нанонауке, нанотехнологии и наноматериалам (Обзор)	6 (1) 9	Ибраев А.Ж., Кульевская Ю.Г., Улезько Г.Г., Галанц Э.А. О реформировании системы подготовки научных кадров. Докторантура PhD (Опыт Казахстана)	5 (1) 15
Тихонов М.Н., Богословский М.М. Факторы интернет-зависимости	6 (2) 1	Маркусова В.А., Рубвальтер Д.А., Либкинд А.Н. Оценки эффективности конкурсного финансирования: обзор библиометрических исследований	5 (1) 23
Берестова Т.Ф. Информационное ресурсосведение как новое научное направление: постановка проблемы	7 (1) 1	Сукиасян Э.Р. Библиотечно-библиографическая классификация (ББК). Динамика развития (конец 2014 г.)	5 (1) 30
Лещёв С.В. Инфогенез и инфотектоника электронной культуры: новые горизонты информационных технологий	7 (1) 10	Сюнтюренко О.В. Перспективы использования интернет-СМИ, журналов открытого доступа и социальных медиа в научно-технической сфере	6 (1) 30
Голицына О.Л., Куприянов В.М., Максимов Н.В. Информационные и технологические решения в задачах управления знаниями	8 (1) 1	Куракова Н.Г., Цветкова Л.А., Ерёмченко О.А. Дисциплинарная структура мировой науки и оптимизация сети диссертационных советов в России: дисбаланс или гармонизация	6 (1) 37
Миролюбова Т.В. Мировой и национальный рынки информационных ресурсов: современные особенности и влияние на экономику	9 (1) 2	Брежнева В.В., Гиляревский Р.С. Еще раз об информационном менеджменте и информатике	7 (1) 15
Сюнтюренко О. В. Сетевые технологии информационного противоборства и манипуляции общественным сознанием	10 (1) 1	Ставинский Е.Н., Романова М.С., Ситникова И.С., Ильина О.И. Оценка деятельности академического института: попытка комплексного подхода	7 (1) 18
Плешкевич Е.А. Документоцентризм как теоретико-методологическая проблема и пути ее решения	11 (1) 1	Мохначева Ю.В. Влияние различных форм соавторства на научную продуктивность российских учёных в области молекулярной биологии	8 (1) 13
Грушо А.А., Забейайло М.И., Зацаринный А.А., Писковский В.О., Борохов С.В. О возможностях приложений интеллектуального анализа данных в задачах обеспечения информационной безопасности облачных сред	11 (2) 1	Мартиросян З.Г., Саркисян Д.Б. Опыт подготовки кадров в сфере долгосрочного сохранения цифровой информации в Республике Армения: Обзор	8 (1) 26
Нестерович Ю.В. Концепты информации и интеллектуального продукта в рамках новой инфолого-документологической парадигмы	12 (1) 1	Арский Ю.М., Никольская И.Ю., Гоннова С.М. Формирование системы тематической классификации с целью развития информационного обмена в научно-технической сфере	9 (1) 23
Еркимбаев А.О., Зицерман В.Ю., Кобзев Г.А., Косинов А.В. Связывание онтологий с базами данных по свойствам веществ и материалов	12 (2) 1	Громова Е.В., Денисов А.С., Кочетков А.В. Организация информационной работы библиотеки технического вуза с учетом индикаторов ее web-сайта	9 (1) 27
ОРГАНИЗАЦИЯ ИНФОРМАЦИОННОЙ РАБОТЫ			
Антопольский А.Б. Алгоритм и процедуры определения уровня развития информационных ресурсов в научных и образовательных организациях	1 (1) 14	Еркимбаев А.О., Зицерман В.Ю., Кобзев Г.А., Трахтенгерц М.С. Универсальная система метаданных для характеристики наноматериалов	10 (1) 8

Шемберко Л.В., Слива А.И. Библиографическая база данных по социологии: принципы и инструменты формирования, структура, особенности поиска	10 (1) 21	Трусова Ю. О., Белоозеров В. Н. Представление классификационных систем в виде онтологий (Обзор)	11 (1) 34
Ширяев А.А. Управленческие информационные системы в научной сфере	10 (1) 31	Буйлова Н.М., Зицерман В.Ю., Кобзев Г.А. Солнечные элементы и нанотехнологии: библиометрический анализ публикаций, отраженных в РЖ «Физика» ВИНТИ РАН и базе данных Web of Science	12 (1) 33
Саркисян Д.Б., Гончарова Ю.Г., Гуров А.Н., Марков И.А. Деятельность мировой системы данных по обеспечению всеобщего и равноправного доступа к научным данным	11 (1) 8	Баранов Д.А., Белоусов К.И., Зелянская Н.Л., Карлина Т.В. Об использовании экономических классификаторов для индексации научных публикаций	12 (1) 40
Либкинд А.Н., Терехов А.И., Маркусова В.А., Рубвальтер Д.А., Либкинд И.А. Результаты выполнения конкурсных исследовательских проектов: библиометрия вклада различных групп ученых, организаций, городов, регионов и стран	11 (1) 16	ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ	
Буцык С.В. О сравнительных параметрах уровня информатизации российских вузов в современный период	11 (1) 29	Забежайло М.И. К вопросу о достаточности оснований для принятия результатов интеллектуального анализа данных средствами ДСМ-метода	1 (2) 1
Гуськов А.Е., Быховцев Е.С., Косяков Д.В. Альтернативная вебметрика: исследование веб-трафика сайтов научных организаций	12 (1) 12	Клещев А.С., Черняховская М.Ю., Шалфеева Е.А. Особенности автоматизации интеллектуальной деятельности	1 (2) 10
Варганова Г.В., Мутьев В.А. Профессионализация студентов библиотечно-информационных факультетов вузов культуры	12 (1) 29	Виноградов Д.В. Вероятность порождения случайного ДСМ-сходства при наличии контр-примеров	3 (2) 1
ДОКУМЕНТАЛЬНЫЕ ИСТОЧНИКИ ИНФОРМАЦИИ		Харчевникова Н.В., Блинова В.Г., Добрынин Д.А., Журков В.С. Использование интеллектуальной ДСМ-системы для анализа связи структуры нитрозамещенных бензолов с их мутагенной активностью в тесте Эймса	3 (2) 6
Демидов Д.Д. Реферативные журналы как источник информации по агроинженерии	2 (1) 20	Гусакова С.М. Корректность ДСМ-рассуждений для однородных стратегий	5 (2) 1
Родионов И.И., Цветкова В.А., Родина И.В., Каширская Н.Ф. Распространение научных периодических изданий: состояние и перспективы развития	7 (1) 22	Панкратова Е.С., Добрынин Д.А. Компьютерная интеллектуальная ДСМ-система для прогнозирования послеоперационных осложнений	6 (2) 9
Сергеева В.В. Исследование взаимосвязи относительной публикационной активности страны и структуры публикационного потока	7 (1) 28	Ольшанский Д.Л. Подбор алгоритма для параллельной реализации метода сходства в интеллектуальных ДСМ-системах	7 (2) 1
Панкеев И.А. Научное произведение в аспекте авторского права: законодательные новеллы	8 (1) 31	Финн В.К. Обнаружение эмпирических закономерностей в последовательностях баз фактов посредством ДСМ-рассуждений	8 (2) 1
Иншакова Н.Г. Информационное сопровождение книжного издания: актуальный взгляд на аппарат книги	8 (1) 35	Ольшанский Д.Л. Информационная среда WebChem-JSM для реализации ДСМ-метода в задачах фармакологии	9 (2) 2
Лукашевич А.В., Лукашевич Н.Л. Обзор используемых в России универсальных классификаций в области исследований Земли из космоса	9 (1) 32	Гусакова С.М., Панкратова Е.С. О корректности ДСМ-рассуждений в медицинских системах	9 (2) 12
Динер Е.В. Книговедческие аспекты электронной книги: опыт регионального исследования	9 (1) 38		

Лыфенко Н.Д. Об одном подходе к классификации текстовых данных, основанном на идеях Д.С.Милля 11 (2) 12

ИНФОРМАЦИОННЫЕ СИСТЕМЫ

Петров В.А., Веселовский А.В., Кузьмина Д.А., Платэ А.Н., Гальберг Т.В. Трехмерное пространственно-временное ГИС-моделирование 2 (2) 1

Малыгин Е.Н., Егоров С.Я., Шаронин К.А. Алгоритм автоматизированного построения математической модели при компоновке промышленных объектов 2 (2) 7

Нефедов О.М., Королева Л.М., Трепалин С.В., Бессонов Ю.Е., Чуракова Н.И. Разработка интегрированной системы структурной химической информации 12 (2) 17

ИНФОРМАЦИОННЫЙ ПОИСК

Федорец О.В. Автоматический поиск соответствий между системами классификации на примере сопоставления рубрик МПК и ГРНТИ 7 (2) 31

Пименов Е.Н. Лингвистические основания путеводителя по вопросам сохранности документов 9 (2) 24

ИНФОРМАЦИОННЫЕ ЯЗЫКИ

Ольшанский Д.Л., Добрынин Д.А., Блинова В.Г. Автоматическое кодирование химических соединений фрагментарным кодом суперпозиций подструктур 5 (2) 7

ИНФОРМАЦИОННЫЙ АНАЛИЗ

Гнатышак Д. В. Однопроходный алгоритм трикластеризации 2 (2) 16

Громова Е.В., Денисов А.С., Кочетков А.В. Библиометрический анализ журналов, в которых публикуются преподаватели технических университетов 3 (2) 12

Миндели Л.Э., Маркусова В.А. Библиометрические исследования научного сотрудничества: обзор мировых тенденций 4 (2) 6

Булдакова Т.И., Миков Д.А. Методика анализа информационных рисков с применением нейро-нечёткой сети 4 (2) 13

Яцко В.А. Метод автоматической классификации текстов, основанный на законе Ципфа 5 (2) 19

Голицына О.Л., Максимов Н.В. Сравнительный структурно-статистический анализ лексики и связей информационно-поисковых тезаурусов 6 (2) 14

Цисун Е, Шелов С.Д. О классификации номенов и номенклатурных наименований (на материале наименований товаров) 6 (2) 29

Ежела В.В., Клименко С.В., Райков А.Н., Шарнин М.М. Семантический подход к оценке качества научных публикаций 7 (2) 13

Петрин А.А. Обработка информации с использованием естественных и искусственных нейронных сетей: сравнительный анализ 7 (2) 19

Московкин В.М., Голиков Н.А., Серкина О.В. Конструирование метрик цитирования нового поколения 8 (2) 30

Когаловский М.Р., Паринов С.И. Таксономия семантических связей информационных объектов контента научной электронной библиотеки 9 (2) 15

Грибова В.В., Клещев А.С., Москаленко Ф.М., Тимченко В.А. Двухуровневая модель сложноструктурированных информационных единиц, соответствующая метафоре анкетирования 10 (2) 1

Ключенко Т.И. Методологическая ценность «библиотечной математики» в междисциплинарном дискурсе 11 (2) 24

Грибова В.В., Клещев А.С., Москаленко Ф.М., Тимченко В.А. Модель порождения орграфов информации по орграфу метаинформации для двухуровневой модели сложноструктурированных информационных единиц 12 (2) 26

АВТОМАТИЗАЦИЯ ОБРАБОТКИ ТЕКСТА

Кутузов А.Б., Кузьменко Е.А. Использование корпусных технологий для изучения ошибок: *learner corpora* на факультете филологии НИУ ВШЭ 1 (2) 21

Тагабилева М. Г. О некоторых нестандартных случаях реализации модели образования композитов со значением *nomina agentis* в русском языке 1 (2) 27

Архангельский Т.А., Созинова О.А. Мультимедийный корпус языка идиш 3 (2) 18

Якушевич И. В. Знаково-семантическая структура поэтического символа и символа-симулякра 3 (2) 25

Хайруллин В.И. Культурно-прагматическая адаптация при переводе как средство извлечения информации 3 (2) 30

Тао Ю., Захаров В.П. Разработка и использование параллельного корпуса русского и китайского языков 4 (2) 18

Кочеткова Н.А. Метод извлечения технических терминов с использованием усовершенствованной меры странности	5 (2) 25	Петрина А.М. О взаимодействии интеллектуальных роботов, программного обеспечения и компьютеров	2 (2) 31
Падучева Е.В. Коммуникативная структура предложения	10 (2) 11	Плющ М.А. Поиск в Интернете сведений о шести тысячах книг в библиотеке Д.М. Голицына: происхождение легенды и определение возможного размера библиотеки	4 (1) 34
Гроховский П.Л., Захаров В.П., Смирнова М.О., Хохлова М.В. Корпус памятников тибетской грамматической традиции	10 (2) 26	Дружинина Е.Г. О конференции в Санкт-Петербургском государственном институте культуры	5 (1) 34
Большакова Е.И., Большаков И.А. Аффиксальный критерий паронимии для построения компьютерного словаря паронимов русского языка	11 (2) 28	Сковородина И.С., Ефремов П.В. О конференции «Технологии создания, агрегации и использования научного и образовательного контента»: от концепции к проектам и практическим решениям	5 (1) 35
Розина Р.И. Семантические переходы в разговорной речи	11 (2) 36	Плешкевич Е.А. На пути к информационной картине мира Джеймса Глика [Рец. на кн.]	5 (2) 33
СПРАВОЧНО-ИНФОРМАЦИОННЫЙ РАЗДЕЛ			
Швецова-Водка Г.Н. Документ в теории документальной информации Е.А. Плешкевича [Рец. на кн.]	1 (1) 27	Арутюнов В.В. Об итогах III международной научно-практической конференции «Современные информационные технологии в профессиональной деятельности» (СИТ-2014)	2 (1) 27
Арутюнов В.В. О международной научно-практической конференции «Современные информационные технологии в профессиональной деятельности» (СИТ-2014)	2 (1) 27	Арутюнов В.В. Об итогах III международной научно-практической конференции «Современные проблемы и задачи обеспечения информационной безопасности»	7 (1) 34
		Гиляревский Р.С. Своевременная книга по наукометрии [Рец. на кн.]	9 (1) 43

Авторский указатель

Антопольский А.Б.	1 (1) 14	Блинова В.Г.	3 (2) 6	Голиков Н.А.	8 (2) 30
	3 (1) 3		5 (2) 7	Голицына О.Л.	6 (2) 14
Антошкова О.А.	3 (1) 19	Богословский М.М.	6 (2) 1		8 (1) 1
	3 (1) 37	Большаков И.А.	11 (2) 28	Гоннова С.М.	3 (1) 57
	3 (1) 43	Большакова Е.И.	11 (2) 28		9 (1) 23
	3 (1) 57	Борохов С.В.	11 (2) 1	Гончарова Ю.Г.	11 (1) 8
Арский Ю.М.	9 (1) 23	Брежнева В.В.	5 (1) 7	Грибова В.В.	10 (2) 1
Арутюнов В.В.	2 (1) 27		7 (1) 15		12 (2) 26
	7 (1) 34	Буйлова Н.М.	12 (1) 33	Громова Е.В.	1 (1) 25
Архангельский Т.А.	3 (2) 18	Булдакова Т.И.	4 (2) 13		3 (2) 12
Астахова Т.С.	3 (1) 19	Буцык С.В.	11 (1) 29		9 (1) 27
	3 (1) 37	Быховцев Е.С.	12 (1) 12	Гроховский П.Л.	10 (2) 26
Астахова Л.В.	4 (1) 1			Грушо А.А.	11 (2) 1
		Варганова Г.В.	12 (1) 29	Гуреев В.Н.	2 (1) 8
Баранов Д.А.	12 (1) 40	Верная Л. А.	3 (1) 70	Гуров А.Н.	11 (1) 8
Белоозеров В.Н.	3 (1) 3	Веселовский А.В.	2 (2) 1	Гусакова С.М.	5 (2) 1
	3 (1) 47	Виноградов Д.В.	3 (2) 1		9 (2) 12
	3 (1) 60	Галанц Э.А.	5 (1) 15	Гуськов А.Е.	12 (1) 12
	11 (1) 34	Гальберг Т.В.	2 (2) 1		
Белоусов К.И.	12 (1) 40	Гиляревский Р.С.	5 (1) 7	Двоеносова Г.А.	6 (1) 1
Берестова Т.Ф.	7 (1) 1		7 (1) 15	Демидов Д.Д.	2 (1) 20
Бессонов Ю.Е.	12 (2) 17		9 (1) 43	Денисов А.С.	3 (2) 12
		Гнатышак Д. В.	2 (2) 16		9 (1) 27

Динер Е.В.	9 (1) 38	Кочеткова Н.А.	5 (2) 25	Плешкевич Е.А.	4 (1) 14
Дмитриева Е.Ю.	3 (1) 3	Кузьменко Е.А.	1 (2) 21		5 (2) 33
	3 (1) 23	Кузьмина Д.А.	2 (2) 1		11 (1) 1
	3 (1) 37	Кульевская Ю.Г.	5 (1) 15	Плющ М.А.	4 (1) 34
Добрынин Д.А.	3 (2) 6	Куприянов В.М.	8 (1) 1		
	5 (2) 7	Кутузов А.Б.	1 (2) 21		
	6 (2) 9	Куракова Н.Г.	6 (1) 37	Раевская Е.Г.	4 (1) 24
Дружинина Е.Г.	5 (1) 34			Райков А.Н.	7 (2) 13
		Лешёв С.В.	7 (1) 10	Родина И.В.	7 (1) 22
Егоров С.Я.	2 (2) 7	Либкинд А.Н.	5 (1) 23	Родионов И.И.	5 (1) 10
Ежела В.В.	7 (2) 13		11 (1) 16		7 (1) 22
Ерёмченко О.А.	6 (1) 37	Либкинд И.А.	11 (1) 16	Розина Р.И.	11 (2) 36
Еркимбаев А.О.	10 (1) 8	Лукашевич А. В.	9 (1) 32	Романова М.С.	7 (1) 18
	12 (2) 1	Лукашевич Н. Л.	9 (1) 32	Рубвальтер Д.А.	5 (1) 23
Ермолаева А.В.	5 (1) 1	Лыфенко Н.Д.	11 (2) 12		11 (1) 16
Ермолаева В.В.	1 (1) 25				
Ефременкова В.М.	3 (1) 27	Мазов Н.А.	2 (1) 8	Саркисян Д.Б.	3 (1) 19
Ефремов П.В.	5 (1) 35	Максимов Н.В.	6 (2) 14		8 (1) 26
			8 (1) 1		11 (1) 8
Журков В.С.	3 (2) 6	Малыгин Е.Н.	2 (2) 7	Семенюк Э.П.	1 (1) 1
		Маркарова Т.С.	3 (1) 3	Сергеева В.В.	7 (1) 28
Забежайло М.И.	1 (2) 1	Марков И.А.	11 (1) 8	Серкина О.В.	8 (2) 30
	11 (2) 1	Маркусова В.А.	4 (2) 6	Ситникова И.С.	7 (1) 18
Захаров В.П.	4 (2) 18		5 (1) 23	Сковородина И.С.	5 (1) 35
	10 (2) 26		11 (1) 16	Слива А.И.	10 (1) 21
Зацаринный А.А.	11 (2) 1	Мартиросян З.Г.	8 (1) 26	Смирнова М.О.	10 (2) 26
Зелянская Н.Л.	12 (1) 40	Миков Д.А.	4 (2) 13	Смирнова О.В.	3 (1) 27
Зибарева И.В.	6 (1) 9	Миндели Л.Э.	4 (2) 6		3 (1) 35
Зицерман В.Ю.	10 (1) 8	Миролюбова Т.В.	9 (1) 2		3 (1) 37
	12 (1) 33	Москаленко Ф.М.	10 (2) 1		3 (1) 40
	12 (2) 1		12 (2) 26	Смылова И.С.	3 (1) 32
		Московкин В.М.	8 (2) 30	Созинова О.А.	3 (2) 18
Ибраев А.Ж.	5 (1) 15	Мохначева Ю.В.	8 (1) 13	Соловьёва И.М.	3 (1) 23
Ивановский А.А.	3 (1) 52	Мутьев В.А.	12 (1) 29		3 (1) 32
Ильина О.И.	7 (1) 18			Ставинский Е.Н.	7 (1) 18
Иншакова Н.Г.	8 (1) 35			Сукиасян Э.Р.	5 (1) 30
		Нестерович Ю.В.	12 (1) 1	Сурикова Н.Г.	3 (1) 37
		Нефедов О.М.	12 (2) 17	Сюнтюренко О.В.	2 (1) 1
		Никольская И. Ю.	3 (1) 57		6 (1) 30
			9 (1) 23		10 (1) 1
Калёнов Н. Е.	3 (1) 60				
	4 (2) 1	Ольшанский Д.Л.	5 (2) 7	Тагабилева М. Г.	1 (2) 27
Карлина Т.В.	12 (1) 40		7 (2) 1	Тао Ю.	4 (2) 18
Карпюк Д.В.	4 (1) 19		9 (2) 2	Терехов А.И.	11 (1) 16
Каширская Н.Ф.	7 (1) 22			Тимченко В.А.	10 (2) 1
Клещев А.С.	1 (2) 10				12 (2) 26
	10 (2) 1	Падучева Е.В.	10 (2) 11	Тихонов М.Н.	6 (2) 1
	12 (2) 26	Панкеев И.А.	8 (1) 31	Трахтенгерц М.С.	10 (1) 8
Клименко С.В.	7 (2) 13	Панкратова Е.С.	6 (2) 9	Трепалин С.В.	12 (2) 17
Ключенко Т.И.	11 (2) 24		9 (2) 12	Трусова Ю. О.	11 (1) 34
Кобзев Г.А.	10 (1) 8	Паринов С.И.	9 (2) 15		
	12 (1) 33	Петрин А.А.	7 (2) 19	Улезько Г.Г.	5 (1) 15
	12 (2) 1	Петрина А.М.	2 (2) 31		
Когаловский М.Р.	9 (2) 15	Петров В.А.	2 (2) 1	Федорец О.В.	7 (2) 31
Королева Л.М.	12 (2) 17	Пименов Е.Н.	9 (2) 24	Финн В.К.	8 (2) 1
Косинов А.В.	12 (2) 1	Писковский В.О.	11 (2) 1		
Косяков Д.В.	12 (1) 12	Платэ А.Н.	2 (2) 1		
Кочетков А.В.	1 (1) 25				
	3 (2) 12				
	9 (1) 27				
				Хайруллин В.И.	3 (2) 30

Харчевникова Н.В.	3 (2) 6	Чуракова Н.И.	12 (2) 17	Шелов С.Д.	6 (2) 29
Хохлова М.В.	10 (2) 26			Шемберко Л.В.	10 (1) 21
Цветкова В.А.	5 (1) 10	Шабурова Н.Н.	3 (1) 47	Ширяев А.А.	10 (1) 31
	7 (1) 22		3 (1) 55		
Цветкова Л.А.	6 (1) 37	Шалфеева Е.А.	1 (2) 10	Якушевич И. В.	3 (2) 25
Цисун Е.	6 (2) 29	Шарнин М.М.	7 (2) 13	Яцко В.А.	5 (2) 19
Черняховская М.Ю.	1 (2) 10	Шаронин К.А.	2 (2) 7		
		Швецова-Водка Г.Н.	1 (1) 27		

**Федеральное государственное бюджетное учреждение науки
ВСЕРОССИЙСКИЙ ИНСТИТУТ НАУЧНОЙ И ТЕХНИЧЕСКОЙ
ИНФОРМАЦИИ РОССИЙСКОЙ АКАДЕМИИ НАУК**

предлагает научным работникам, аспирантам и другим специалистам в области естественных, точных и технических наук, желающим быстро и эффективно опубликовать результаты своей научной и научно-производственной деятельности, использовать способ публикации своих работ через систему депонирования.

«Депонирование (передача на хранение) – особый метод публикации научных работ (отдельных статей, обзоров, монографий, сборников научных трудов, материалов научных конференций, симпозиумов, съездов, семинаров) узкоспециального профиля, разрешенных в установленном порядке к открытому опубликованию, широкое тиражирование которых, как правило, в силу их узкой специализации, не считается целесообразным, а также работ широкого профиля, срочная информация о которых необходима для утверждения их приоритета. Депонирование предусматривает прием, учет, регистрацию, хранение научных работ и обязательное размещение информации о них в специальных информационных изданиях».

Подготовка и передача на депонирование научных работ происходит в соответствии с «Инструкцией о порядке депонирования научных работ по естественным, техническим, социальным и гуманитарным наукам» (М., 2013).

Депонированные научные работы находятся на хранении в депозитарии ВИНТИ РАН, копии работ предоставляются заинтересованным организациям и специалистам на бумажном и электронном носителях и являются официальной публикацией.

Информация о депонированных научных работах включается в информационные издания ВИНТИ РАН, в РЖ ВИНТИ РАН и БД ВИНТИ РАН и аннотированный библиографический указатель «Депонированные научные работы».

Подать научную работу на депонирование можно, обратившись в Отдел депонирования ВИНТИ РАН по адресу:

125190, Москва, ул. Усиевича, 20.

ВИНТИ РАН, Отдел депонирования научных работ.

Тел.: 8 (499) 155-43-28, Факс: 8 (499) 943-00-60.

e-mail: dep@viniti.ru

С инструкцией о порядке депонирования можно ознакомиться на сайте ВИНТИ РАН: <http://www.viniti.ru>

Центр (Отдел) научно-информационного обслуживания (ЦНИО) ВИНТИ РАН

предлагает услуги по предоставлению информационно-аналитических обзоров

ВИНТИ РАН осуществляет подготовку информационно-аналитических обзоров по инновационным и приоритетным направлениям научных исследований в области точных, естественных и технических наук. Обзоры готовятся ведущими специалистами ВИНТИ, работающими в определенных областях науки и техники. Аналитические материалы содержат результаты анализа и обобщения информации по актуальным научным проблемам, а в некоторых случаях – и прогностические выводы. Основой для составления обзоров служит отечественная и зарубежная научно-техническая литература, доступная ВИНТИ РАН: фонд НТЛ, включающий более 2 млн отечественных и иностранных журналов, книг, депонированных рукописей, авторефератов диссертаций и другой научной литературы, ретроспектива – с 1987 года. Имеется доступ к базам данных и Интернет-ресурсам: БД ВИНТИ (разработка ВИНТИ), БД SCOPUS, БД зарубежных патентов и другим. Кроме того, ВИНТИ доступны зарубежные электронные платформы ряда ведущих научных издательств, выпускающих основную часть академических рецензируемых журналов, в полнотекстовом варианте.

Основные тематические направления предлагаемых обзоров:

- Науки о жизни;
- Физико-математические науки;
- Химия и науки о материалах;
- Индустрия наносистем и материалов;
- Науки о Земле;
- Рациональное природопользование;
- Информационно-телекоммуникационные системы;
- Энергетика, энергоэффективность, энергосбережение;
- Транспортные, авиационные и космические системы;
- Производственные технологии.

Предлагается подготовка и заказ информационно-аналитических обзоров и материалов по тематике заказчика. Такие обзоры могут относиться к упомянутым выше тематическим направлениям, но могут иметь и междисциплинарный характер. В этом случае обзоры отражают актуальную научную информацию и научные достижения, происходящие на стыке наук.

Более подробная информация о приобретении, заказе и цене обзоров представлена на сайте ВИНТИ www.viniti.ru

Приобретение и заказ обзоров от юридических лиц проводится на договорной основе. Форма договора для последующего оформления представлена на сайте ВИНТИ.

Оформление договоров и других необходимых документов производится Центром научно-информационного обслуживания ВИНТИ (ЦНИО). Возможен прием заказов от физических лиц, оплата производится на расчетный счет или в кассу ВИНТИ РАН.

Выполненные в ВИНТИ обзоры предоставляются заказчикам в печатном виде либо в электронном варианте после оплаты заказа.

Обращаться в ЦНИО ВИНТИ:

- адрес: 125190, Россия, г. Москва, ул. Усиевича, 20.
- телефоны: 8(499) 155 -42 -43, 8(499) 155 -42 -17
- эл. почта cnio@viniti.ru, fdk@viniti.ru.
- факс 8(499) 930 -60 -00 (для ЦНИО).

УВАЖАЕМЫЕ КОЛЛЕГИ!

ВИНИТИ РАН предлагает Вашему вниманию Реферативный Журнал в электронной форме

РЖ в электронной форме (ЭлРЖ) выпускается по всем разделам естественных, технических и точных наук.

Каждый номер ЭлРЖ является полным аналогом печатного номера РЖ по составу описаний документов, их оформлению и расположению. Он сопровождается оглавлением, указателями.

ЭлРЖ представляет собой информационную систему, снабженную поисковым аппаратом и позволяющую пользователю на персональном компьютере:

- читать номер РЖ, последовательно листая рефераты;
- просматривать рефераты отдельных разделов по оглавлению;
- обращаться к рефератам по указателям авторов, источников, ключевых слов;
- проводить поиск документов по словам и словосочетаниям;
- выводить текст описаний документов во внешний файл.

ЭлРЖ в версии Windows Вы можете получить за текущий год с любого номера, а также за предыдущие годы.

Подробную информацию Вы можете получить:

Адрес: 125190, Россия, Москва, ул. Усиевича, 20, ВИНТИ РАН

Телефон: 8 (499) 155-46-20

Телефон/Факс: 8 (499) 155-45-25

E-mail: zinovyeva@viniti.ru, davydova@viniti.ru